AWARD NUMBER: W81XWH-12-1-0041

TITLE: Understanding and Targeting Epigenetic Alterations in Acquired Bone Marrow Failure

PRINCIPAL INVESTIGATOR: Omar Abdel-Wahab

CONTRACTING ORGANIZATION: Memorial Sloan-Kettering Cancer Center
New York, NY 10065

REPORT DATE: May 2015

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
May 2015	Annual Summary	1May2014 - 30Apr2015

4. TITLE AND SUBTITLE
Understanding and Targeting Epigenetic Alterations in Acquired Bone Marrow Failure

5a. CONTRACT NUMBER
W81XWH-12-1-0041

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Omar Abdel-Wahab

E-Mail: abdelwao@mskcc.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Memorial Sloan-Kettering Cancer Center
New York, NY 10065

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Systematic genomic discovery efforts in patients with bone marrow failure due to myelodysplastic syndrome (MDS) has led to the rapid discovery of recurrent somatic genetic alterations underlying these disorders. Remarkably, a large number of these mutations occur in genes whose function is known, or suspected, to be involved in epigenetic regulation of gene transcription. This includes mutations in *ASXL1, TET2,* and *EZH2.* The goals of our proposal were to (1) perform functional genetic characterization of these alterations, (2) determine if these alterations are therapeutically targetable, and (3) perform detailed genomic analysis of specific subsets of MDS patients with no known genetic alterations and with severe bone marrow failure to discover additional genetic alterations contributing to MDS pathogenesis. Since funding of this award we have made major progress in (1) understanding the impact of *ASXL1* mutations and loss on chromatin (Abdel-Wahab, *et al. Cancer Cell* 2012), (2) identifying the *in vivo* biological effects of deletion of *Asxl1* and *Tet2* alone and in combination with one another (Abdel-Wahab, *et al. J Exp Med* 2013), and (3) identified the genome-wide effects of Asxl1 on transcription (Abdel-Wahab, *et al. J Exp Med* 2013 and *Abdel-Wahab, O,* et al. *Leukemia* 2013). More recently we have identified that recently common mutations in the splicing machinery in MDS also may impact the function of these epigenetic modifiers (Kim, E, *et al. Cancer Cell* 2015).

15. SUBJECT TERMS
ASXL1; Bone marrow failure; Myelodysplastic Syndrome; TET2

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** USAMRMC
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	40	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Introduction

Increasing use of genomic discovery efforts in patients with bone marrow failure due to myelodysplastic syndrome (MDS) has led to the rapid discovery of a series of recurrent genetic abnormalities underlying these disorders. Remarkably, a large number of these alterations appear to be in genes whose function is known, or suspected, to be involved in epigenetic regulation of gene transcription. In the last 3 years alone, mutations in the genes *TET2*, *ASXL1*, *DNMT3a*, and *EZH2* have all been found to be frequent mutations amongst patients with MDS. Mutations in several of these genes have proven to be important markers of disease outcome with *ASXL1* and *EZH2* mutations recurrently being identified as adverse prognosticators in MDS patients. Identification of frequent mutations in epigenetic modifiers has also highlighted the fact that a number of these genes encode enzymes and/or result in alterations in enzymatic alterations which may represent novel, tractable therapeutic targets for MDS patients. In this proposal, we originally aimed to identify (a) if mice with genetically engineered deletion of epigenetic modifiers mutated in MDS would serve as valuable murine models of MDS, (b) if mutations in epigenetic modifiers may specifically impact DNA methylation and/or histone post-translational modifications in a manner that is therapeutically targetable, and (c) if additional mutations must exist in patients with specific subsets of MDS with the worst clinical outcome. Since awarding of the proposal, we have made major insights into the epigenomic function of ASXL1 as well as the biological impact of conditional deletion of *Asxl1* alone and in combination with other genetic alterations including *Tet2* deletions and NRasG12D overexpression. In addition, we have recently identified that an additional class of very frequency mutations in MDS patients affecting the spliceosome impacts EZH2 function. This work has resulted in several publications, multiple oral presentations at national meetings, and has been used as the basis for several additional foundation awards (from Damon Runyon Foundation, the V Foundation, and the Evans Foundation) and is the basis for an NIH R01 application I have pending.

Keywords:

5-azacytidine, ASXL1, Decitabine, Epigenetics, EZH2, Genomics, Mouse models, Myelodysplastic Syndromes, Splicing, SRSF2, TET2.

Accomplishments

Key Research Accomplishments

- Developed and published the first conditional knockout mouse for *Asxl1* as well as the first murine model with combined *Asxl1* and *Tet2* deletion. We believe these models are valuable genetically accurate murine models of acquired bone marrow failure.
- Identified the biological effects of *Asxl1* loss on hematopoiesis, alone and in combination with other co-occurring genetic alterations.
- Generated the first murine model of spliceosomal mutations as seen in patients with MDS.
- Identified an important intersection of spliceosomal gene alterations on the epigenome of MDS.

In addition to the above summary, below is a more detailed summary of accomplishments organized by Tasks from the original grant submission:

Task 1. "Obtain DoD ACURO approval for the use of animals in the experiments outlined below in Tasks 2 to 4."

We have nearly completed DoD ACURO approval for all experiments related to this award. We are awaiting final confirmation on approval from DoD currently.

Task 2. "Complete characterization of mice with conditional deletion of Asxl1 alone and Asxl1 combined with Tet2 (Months 1-24) at the work performance site of Memorial Sloan-Kettering Cancer Center."

As noted in our annual review in 2014, we completed generation of mice with deletion of *Asxl1, Tet2,* or both using multiple different Cre recombinases. This work was recently published in 2013 in the *Journal of Experimental Medicine* (**Abdel-Wahab, O**, *et al*. *J Exp Med* 2013 Nov 18;210(12):2641-59) and have been used by the MDS research community internationally. We have deposited these mice at the Jackson Laboratory for public use.

Task 3. Continue development of mice with Ezh2 deletion alone and characterize mice with compound deletion of Ezh2/Tet2 and Ezh2/Asxl1 (Months 1-24) at the work performance site of Memorial Sloan-Kettering Cancer Center.

We recently generated mice with Ezh2 deletion in the postnatal compartment (*Mx1-cre Ezh2fl/fl*) mice and mice with compound deletion of *Ezh2* and *Asxl1.* From these murine models we have identified that:

(i) Hematopoietic stem cells (HSCs) from mice with compound *Asxl1/Ezh2* loss have impaired self-renewal compared with HSCs from littermate control mice as well as mice with deletion of either gene alone.

(ii) A high proportion of wildtype mice reconstituted with bone marrow from mice with compound *Asxl1/Ezh2* (Mx1-cre Asxl1fl/fl Ezh2fl/fl) deletion die of bone marrow failure within weeks of deletion of these genes. Surviving mice are characterized by anemia and leukopenia as well as morphologic dysplasia.

The above phenotypes of mice with compound deletion of both Asxl1 and Ezh2 are dramatic and we are now working to functionally understand the mechanism by which deletion of these 2 genes impairs HSC function.

In addition to the above, we have recently identified the unexpected observation that mutations in the spliceosomal protein SRSF2, commonly identified in MDS patients, results in mis-splicing of *EZH2.* Interestingly, *SRSF2* mutations and loss-of-function EZH2 mutations in MDS are 100% mutually exclusive but the functional basis for this interaction was not known previously. Our work provided the basis for this observation and identified another mechanism by which EZH2 is dysregulated in MDS. These data were recently published in the following manuscript (see **Appendix #1**):

Kim E, Ilagan JO, Liang Y, Daubner GM, Lee SC, Ramakrishnan A, Li Y, Chung YR, Micol JB, Murphy ME, Cho H, Kim MK, Zebari AS, Aumann S, Park CY, Buonamici S, Smith PG, Deeg HJ, Lobry C, Aifantis I, Modis Y, Allain FH, Halene S, Bradley RK, **Abdel-Wahab O**. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell. 2015 May 11;27(5):617-30. doi:

10.1016/j.ccell.2015.04.006. PubMed PMID: 25965569; PubMed Central PMCID: PMC4429920.

Task 4. Determine the epigenetic contribution of Asxl1 and Ezh2 loss to bone marrow failure through Chromatin immunoprecipitation (ChIP) of histone H3 lysine 27 trimethyl (H3K27me3) followed by next-generation sequencing in primary murine hematopoietic cells (Months 1-24) at the work performance site of Memorial Sloan-Kettering Cancer Center.

As noted in 2 prior annual reports, we have completed detailed characterization of the effects of *ASXL1* mutations and loss using cell lines and primary cells from knockout mice. These results have been published now in 2 papers (Abdel-Wahab, O, *et al. Cancer Cell* 2012 and Abdel-Wahab, O, *et al. J Exp Med* 2013).

Task 5: Determine the effect of Tet2, Asxl1, and Ezh2 loss to a panel of currently clinically utilized compounds in patients with MDS. Drug panel will include decitabine, 5-azacytidine, lenalidomide, cytarabine, daunorubicin, HDACi (vorinostat, romidepsin, panobinostat, AR-42, trichostatin A), HSP-90 inhibitors (AUY-922, PUH-71), and parthenolide (Months 1-24) at the work performance site of Memorial Sloan-Kettering Cancer Center.

We are now performing these experiments *ex vivo* through use of methylcellulose colony assays. In brief, hematopoietic stem/progenitor cells (HSPCs; lineage-negative Sca1+ c-KIT+ cells) from Tet2 knockout, Asxl1 knockout, Ezh2 knockout, and Tet2/Asxl1 double knockout mice are being plated in methylcellulose with a variety of the above compounds for 7 days. We are evaluating the effects of these compounds on restoring colony formation (for Asxl1 and Ezh2 knockout HSPCs) or reducing colony formation (for Tet2 and Tet2/Asxl1 knockout HSPCs). This work is underway.

In addition to the above preclinical experiments, we have recently completed a study analyzing the impact of (i) common mutations in MDS and (ii) patterns of DNA genome-wide methylation on response to decitabine treatment. This was performed on a uniformly treated cohort of 40 patients. Although we did not find any association between mutations and response to decitabine, using the methylation profiles, we developed an epigenetic classifier that accurately predicted DAC response at the time of diagnosis. This work was recently published as follows (see **Appendix #2**):

Meldi K, Qin T, Buchi F, Droin N, Sotzen J, Micol JB, Selimoglu-Buet D, Masala E, Allione B, Gioia D, Poloni A, Lunghi M, Solary E, Abdel-Wahab O, Santini V, Figueroa ME. Specific molecular signatures predict decitabine response in chronic myelomonocytic leukemia. J Clin Invest. 2015 May;125(5):1857-72. doi: 10.1172/JCI78752. Epub 2015 Mar 30. PubMed PMID: 25822018.

Task 5: Perform candidate gene and exome sequencing on DNA samples from 20 MDS patients with *ASXL1* mutations alone (Months 1-6) at the work performance site of Memorial Sloan-Kettering Cancer Center.

In order to complete this task and to inform task #5, we recently performed targeted DNA sequencing on pretreatment DNA samples from a cohort of MDS patients uniformly treated with decitabine. This work, performed in collaboration with MDS clinical expert Dr. Valeria Santini, revealed that *ASXL1* mutations frequently co-occur with mutations in the spliceosome-associated protein *SRSF2* in patients with MDS/MPN overlap syndromes. This interesting finding suggests an interaction by mutations in the epigenome with mutations in the spliceosome. Moreover, this work has resulted in one recent publication as noted above (in “Task 5”).

Task 6: Perform candidate gene and exome sequencing on DNA samples from 40 patients with MDS accompanied by moderate to severe bone marrow fibrosis (Months 1-6) at the work performance site of Memorial Sloan-Kettering Cancer Center.

We have now collected samples from 40 such patients with MDS with bone marrow fibrosis and hope to begin performing DNA sequencing soon. We recently helped to generate a DNA next-generation sequencing panel of 300 genes implicated in cancer pathogenesis at our institution. We will apply this sequencing platform to these MDS samples with the hopes of characterizing any novel mutations associated with this unique subtype of MDS.

Task 7: Present findings at national meetings and publish in peer-reviewed journals (Month 6-36).

I have given 15 presentations at national/international meetings on the work performed with funding from this award in the last year (see list of presentations in **Products** below).

I have also been invited to write several reviews related to the work described in this proposal in well-respected journals including *Journal of Clinical Investigation* (cited in **Products** below).

Impact

Genomic discovery efforts in patients with MDS have revealed that the most frequent somatic mutations in these disorders are in genes involved in either epigenetic regulation or RNA splicing. We and others have recently shown that mutations in the Polycomb-associated gene *ASXL1* and the spliceosmal gene *SRSF2* have adverse prognostic importance in patients with all myeloid malignancies including MDS, acute myeloid leukemia (AML), chronic myelomoncytic leukemia (CMML), and primary myelofibrosis. We therefore have focused on understanding the role of these mutations in MDS pathogenesis. In brief, we have identified that the loss-of-function mutations in *ASXL1* as well as the gain-of-function mutations in *SRSF2* both converge on decreased function of the Polycomb Repressive Complex 2 (PRC2). This work has resulted in multiple genetically accurate models of MDS as well as reagents to screen for novel therapeutic targets for *TET2-, ASXL1*- or *SRSF2*-mutant cells.

Changes/Problems

Nothing to report.

Products

Original Manuscripts:

1: Kim E, Ilagan JO, Liang Y, Daubner GM, Lee SC, Ramakrishnan A, Li Y, Chung YR, Micol JB, Murphy ME, Cho H, Kim MK, Zebari AS, Aumann S, Park CY, Buonamici S, Smith PG, Deeg HJ, Lobry C, Aifantis I, Modis Y, Allain FH, Halene S, Bradley RK, **Abdel-Wahab O**. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell. 2015 May 11;27(5):617-30. doi: 10.1016/j.ccell.2015.04.006. PubMed PMID: 25965569; PubMed Central PMCID: PMC4429920.

2: Meldi K, Qin T, Buchi F, Droin N, Sotzen J, Micol JB, Selimoglu-Buet D, Masala E, Allione B, Gioia D, Poloni A, Lunghi M, Solary E, **Abdel-Wahab O**, Santini V, Figueroa ME. Specific molecular signatures predict decitabine response in chronic myelomonocytic leukemia. J Clin Invest. 2015 May;125(5):1857-72. doi: 10.1172/JCI78752. Epub 2015 Mar 30. PubMed PMID: 25822018.

3: Guzman ML, Yang N, Sharma KK, Balys M, Corbett CA, Jordan CT, Becker MW, Steidl U, Abdel-Wahab O, Levine RL, Marcucci G, Roboz GJ, Hassane DC. Selective activity of the histone deacetylase inhibitor AR-42 against leukemia stem cells: a novel potential strategy in acute myelogenous leukemia. Mol Cancer Ther. 2014 Aug;13(8):1979-90. doi: 10.1158/1535-7163.MCT-13-0963. Epub 2014 Jun 16. PubMed PMID: 24934933; PubMed Central PMCID: PMC4383047.

Review Papers:

1: Lee SC, **Abdel-Wahab O**. The mutational landscape of paroxysmal nocturnal hemoglobinuria revealed: new insights into clonal dominance. J Clin Invest. 2014 Oct;124(10):4227-30. doi: 10.1172/JCI77984. Epub 2014 Sep 17. PubMed PMID: 25244089; PubMed Central PMCID: PMC4191026.

2: Aumann S, **Abdel-Wahab O**. Somatic alterations and dysregulation of epigenetic modifiers in cancers. Biochem Biophys Res Commun. 2014 Dec 5;455(1-2):24-34. doi: 10.1016/j.bbrc.2014.08.004. Epub 2014 Aug 9. Review. PubMed PMID: 25111821.

Presentations:

2014 Center for Medical Genetics, Ghent University, Ghent, Belgium
2014 Evans Foundation MDS Summit, Philadelphia, PA
2014 Plenary Session, AACR Hematologic Sessions, Philadelphia, PA
2014 73rd Annual Meeting of the Japanese Cancer Association, Yokohama, Japan
2014 Seminar, Institute of Medical Sciences, University of Tokyo, Tokyo, Japan
2014 Scientific Workshop on Myeloid Development, 56th Annual Meeting of the American Society of Hematology (ASH), San Francisco, California
2014 Oral Session of Basic and Translation Studies in MDS, 56th Annual Meeting of the American Society of Hematology (ASH), San Francisco, California
2015 7th Biennial Workshop on "Clinical Translation of Epigenetics in Cancer Therapy", St. Augustine, Florida
2015 Dept. of Biochemistry Seminar, University of Virginia, Charlottesville, VA.
2015 Research Seminar Series, Boston Children's Hospital, Boston, MA
2015 Indiana University, Wells Center for Pediatric Research Seminar Series
2015 Evans Foundation, MDS Research Summit, Washington D.C.
2015 Lineberger Cancer Center Seminar, UNC Chapel Hill, Chapel Hill, N.C.
2015 Molecular Aspects of Hematology Workshop, Erasmus University, Rotterdam, Netherlands
2015 20th annual meeting of the European Hematology Association (EHA)

Informatics:
We have generated and published multiple new mRNA sequencing (RNA-Seq) datasets as follows:

- Deep RNA-seq analysis of primary MDS patient samples with and without spliceosomal gene mutations for the purpose of identifying novel splice isoforms.

Funding applied for based on this work:
Applied for and successfully received foundation award funding from the V Foundation Scholar Award, the Edward P. Evans Foundation for MDS Research, and the Starr Family Cancer Foundation.

Participants

Name: Omar Abdel-Wahab
Project Role: PI
Research Identifier: none
Nearest Person Month Worked: 12
Contribution to Project: Designed and carried out experiments described with the assistance of Dr. Eunhee Kim and Mr. Young-Rock Chung.

Name: Eunhee Kim
Project Role: Post-doctoral research fellow
Research Identifier: none
Nearest Person Month Worked: 6
Contribution to Project: Designed and carried out experiments described above.
Funding Support: World Cancer Research Foundation

Name: Young-Rock Chung
Project Role: Research technician
Research Identifier: none
Nearest Person Month Worked: 6
Contribution to Project: Assisted Drs. Abdel-Wahab and Kim on murine experiments.
Funding Support: NIH NCI K08 CA160647-04.

Special Reporting Requirements

Nothing to report.

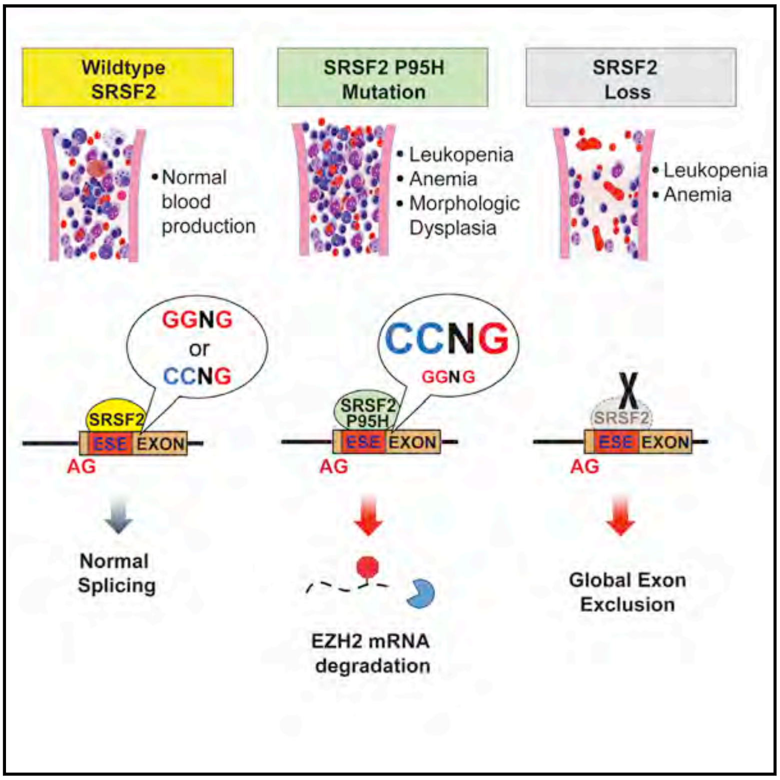
Appendices (please see next page)

Article

Cancer Cell

SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition

Graphical Abstract



Authors

Eunhee Kim, Janine O. Ilagan, ..., Robert K. Bradley, Omar Abdel-Wahab

Correspondence

rbradley@fhcrc.org (R.K.B.), abdelwao@mskcc.org (O.A.-W.)

In Brief

Kim et al. report that myelodysplastic syndrome-associating *SRSF2* mutations alter SRSF2's sequence-specific RNA binding activity, leading to recurrent mis-splicing of key hematopoietic regulators such as *EZH2* and impaired hematopoietic differentiation.

Highlights

- *Srsf2*P95H/wild-type mice develop myelodysplasia but *Srsf2*-deficient mice do not
- Proline 95 mutations change the RNA binding specificity of SRSF2
- Mutant SRSF2 promotes an isoform of *EZH2* that undergoes nonsense-mediated decay
- Restoring EZH2 expression partially rescues hematopoiesis in *Srsf2* mutant cells

Accession Numbers

GSE65349

Kim et al., 2015, Cancer Cell *27*, 617–630
May 11, 2015 ©2015 Elsevier Inc.
http://dx.doi.org/10.1016/j.ccell.2015.04.006

Cancer Cell
Article

SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition

Eunhee Kim,[1,16] Janine O. Ilagan,[2,3,16] Yang Liang,[4,16] Gerrit M. Daubner,[5,16] Stanley C.-W. Lee,[1] Aravind Ramakrishnan,[6,7] Yue Li,[8] Young Rock Chung,[1] Jean-Baptiste Micol,[1] Michele E. Murphy,[6] Hana Cho,[1] Min-Kyung Kim,[1] Ahmad S. Zebari,[2,3] Shlomzion Aumann,[1] Christopher Y. Park,[1,9] Silvia Buonamici,[10] Peter G. Smith,[10] H. Joachim Deeg,[6,7] Camille Lobry,[11,12] Iannis Aifantis,[13] Yorgo Modis,[8,14] Frederic H.-T. Allain,[5] Stephanie Halene,[4,17] Robert K. Bradley,[2,3,17,*] and Omar Abdel-Wahab[1,15,17,*]

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[2]Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[3]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[4]Hematology, Yale Comprehensive Cancer Center and Department of Internal Medicine, Yale University School of Medicine, New Haven, CT 06520, USA
[5]Institute for Molecular Biology and Biophysics, ETH, 8093 Zürich, Switzerland
[6]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[7]Division of Medical Oncology, School of Medicine, University of Washington, Seattle, WA 98109, USA
[8]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA
[9]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[10]H3 Biomedicine, Cambridge, MA 03129, USA
[11]Institut National de la Santé et de la Recherche Medicale (INSERM) U1009, Institut Gustave Roussy, 94805 Villejuif, France
[12]Université Paris-Sud, 91400 Orsay, France
[13]Howard Hughes Medical Institute and Department of Pathology, New York University School of Medicine, New York, NY 10016, USA
[14]Department of Medicine, University of Cambridge, MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK
[15]Leukemia Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[16]Co-first author
[17]Co-senior author
*Correspondence: rbradley@fhcrc.org (R.K.B.), abdelwao@mskcc.org (O.A.-W.)
http://dx.doi.org/10.1016/j.ccell.2015.04.006

SUMMARY

Mutations affecting spliceosomal proteins are the most common mutations in patients with myelodysplastic syndromes (MDS), but their role in MDS pathogenesis has not been delineated. Here we report that mutations affecting the splicing factor SRSF2 directly impair hematopoietic differentiation in vivo, which is not due to SRSF2 loss of function. By contrast, *SRSF2* mutations alter SRSF2's normal sequence-specific RNA binding activity, thereby altering the recognition of specific exonic splicing enhancer motifs to drive recurrent mis-splicing of key hematopoietic regulators. This includes *SRSF2* mutation-dependent splicing of *EZH2*, which triggers nonsense-mediated decay, which, in turn, results in impaired hematopoietic differentiation. These data provide a mechanistic link between a mutant spliceosomal protein, alterations in the splicing of key regulators, and impaired hematopoiesis.

INTRODUCTION

Somatic mutations in genes encoding components of the spliceosome have been identified in a spectrum of human malignancies, including ~60% of patients with myelodysplastic syndromes (MDS) (Bejar et al., 2012; Papaemmanuil et al., 2013; Yoshida et al., 2011). These mutations occur most commonly in *SF3B1* (*Splicing Factor 3b Subunit 1*), *SRSF2* (*Serine/arginine-Rich Splicing Factor 2*), and *U2AF1* (*U2 Small Nuclear RNA Auxiliary Factor 1*) and almost always as

Significance

Frequent somatic mutations affecting components of the spliceosome have been identified in hematologic malignancies; however, the functional role of these mutations is not known. Here we identify that commonly occurring mutations in the spliceosomal gene *SRSF2* impair hematopoietic differentiation and promote myelodysplasia by altering SRSF2's preference for specific exonic splicing enhancer motifs. This results in consistent mis-splicing in a manner that promotes the expression of abnormal isoforms of a number of key hematopoietic regulators, some of which have been linked previously to leukemogenesis (including *BCOR* and *EZH2*). These data provide a mechanistic basis for the enrichment of spliceosomal mutations in myelodysplasia and identify altered RNA recognition as an important driver of leukemogenesis.

heterozygous missense mutations that are mutually exclusive (Papaemmanuil et al., 2011; Wang et al., 2011; Yoshida et al., 2011). Although the genetic data in MDS suggest that these alterations are critical to disease pathogenesis, it remains unknown how these mutations contribute to MDS and whether they are sufficient to induce MDS.

Recent studies have suggested that mutations in the spliceosomal gene *U2AF1* alter RNA splicing (Brooks et al., 2014; Graubert et al., 2012; Ilagan et al., 2015; Przychodzen et al., 2013; Quesada et al., 2012), and studies of gene expression in primary patient samples with and without *U2AF1* mutations have been performed in an effort to identify downstream mis-spliced genes that might contribute to abnormal hematopoiesis (Brooks et al., 2014; Graubert et al., 2012; Ilagan et al., 2015). However, it remains unknown how these mutations contribute to hematopoietic transformation. To date, no studies have investigated the in vivo effects of spliceosomal mutations expressed from the endogenous locus in the correct cellular context, which might allow delineation of how these alleles contribute to MDS pathogenesis.

To test whether spliceosomal gene mutations are sufficient to drive MDS and determine how altered RNA splicing contributes to transformation in vivo, we studied the biological and transcriptional consequences of mutations in *SRSF2*. *SRSF2* mutations occur in 20%–30% of MDS and ~50% of chronic myelomonocytic leukemia (CMML) patients (Papaemmanuil et al., 2013; Yoshida et al., 2011). SRSF2 is a member of the serine/arginine-rich (SR) protein family that contributes to both constitutive and alternative splicing by binding to exonic splicing enhancer (ESE) sequences within pre-mRNA through its RNA recognition motif domain (RRM) (Graveley and Maniatis, 1998; Liu et al., 2000; Schaal and Maniatis, 1999; Zahler et al., 2004). *SRSF2* mutations are consistently associated with adverse outcomes among MDS and acute myeloid leukemia (AML) patients (Papaemmanuil et al., 2013; Vannucchi et al., 2013; Zhang et al., 2012). Despite the clinical importance of *SRSF2* mutations, to date there have been no studies of the functional impact of *SRSF2* mutations on hematopoiesis or splicing. Here we studied the biological and transcriptional effects of somatic expression of the common *SRSF2*P95H mutation in the hematopoietic compartment.

RESULTS

*Srsf2*P95H Mutant Mice Develop MDS, a Phenotype Distinct from Mice with Heterozygous or Homozygous Loss of *Srsf2*

Given the genetic heterogeneity of primary patient samples as well as the fact that stable overexpression of spliceosomal proteins, even in wild-type (WT) form, is poorly tolerated (Lareau et al., 2007), we first generated a murine model for conditional expression of the commonly occurring *SRSF2*P95H mutation from the endogenous murine locus of *Srsf2* (Figure 1A; Figures S1A and S1B). Mice heterozygous for the *Srsf2*P95H allele (*Srsf2*P95H/WT) were crossed to *Mx1-cre* transgenic mice (Kühn et al., 1995) on a C57BL/6 background to allow for inducible expression of Cre recombinase following intraperitoneal injection of polyinosine-polycytosine (pIpC) (12 μg/g every other day for 3 days by injection, as described previously [Moran-Crusio et al., 2011; Figures S1C and S1D; Supplemental Experimental Procedures]). mRNA sequencing (RNA-seq) analysis of hematopoietic stem/progenitor cells (HSPCs) 2 weeks after the last pIpC injection of 6-week-old *Mx1*-cre *Srsf2*P95H/WT and *Mx1*-cre *Srsf2* WT control mice confirmed heterozygous expression of the mutant allele in equal proportion to the remaining WT *Srsf2* allele in *Mx1*-cre *Srsf2*P95H/WT mice (Figure 1B).

It is currently unknown whether the heterozygous *SRSF2*P95H mutation confers a gain of function, haploinsufficient loss of function, or dominant-negative loss of function. We therefore compared expression of the *Srsf2*P95H mutation with the conditional loss of *Srsf2 in vivo* (Wang et al., 2001). Bone marrow (BM) mononuclear cells (MNCs) from 6-week-old CD45.2 *Mx1-cre Srsf2* WT, *Mx1-cre Srsf2*fl/WT (heterozygous floxed mice for inducible deletion of one copy of *Srsf2*), *Mx1-cre Srsf2*fl/fl (homozygous floxed mice for inducible deletion of both copies of *Srsf2*), and *Mx1-cre Srsf2*P95H/WT were transplanted into lethally irradiated congenic CD45.1 recipient mice, followed by pIpC injection 4 weeks later (note that all mice were treated with pIpC to control for any potential phenotypic effects of pIpC administration on biological or splicing phenotypes). This was done to assess for the phenotypic effects of *Srsf2* deletion or mutation in a hematopoietic cell-autonomous manner. Western blot (WB) analysis revealed the deletion of *Srsf2* in BM MNCs from *Mx1-cre Srsf2*fl/fl mice and normal total Srsf2 levels in *Mx1-cre Srsf2*P95H/WT BM MNCs (Figure S1E). Significant leukopenia and anemia were seen in mice with homozygous *Srsf2* deletion or heterozygous expression of the P95H mutation 18 weeks post-transplant (Figures 1C and 1D) that was also seen at earlier time points (Figures S1F and S1G). The presence of similar cytopenias in mice bearing a homozygous *Srsf2* deletion and a heterozygous *Srsf2*P95H point mutation suggested a possible dominant-negative function imposed by the P95H mutation. However, the anemia in *Srsf2*P95H mice was characterized by increased mean corpuscular volume (MCV) of red blood cells relative to WT mice or mice with loss of one to two copies of *Srsf2* (Figure 1E). Moreover, histological assessment of mice 14 weeks post-pIpC revealed prominent BM aplasia in *Srsf2* homozygous knockout (KO) mice, whereas mice expressing the heterozygous P95H mutation had normal BM cellularity (Figure 1F). Platelet counts were normal in *Srsf2*P95H mutant mice at all time points examined (Figure S1H).

Given that macrocytic anemia, a hallmark of anemia in MDS, was present in *Srsf2*P95H mutant mice, we next performed cytological examination of peripheral blood and bone marrow smears from *Mx1-cre Srsf2* WT, *Mx1-cre Srsf2*fl/fl, and *Mx1-cre Srsf2*P95H/WT mice to assess for morphologic dysplasia. This revealed prominent myeloid and erythroid dysplasia in *Srsf2*P95H mice but not in *Mx1-cre Srsf2* WT or *Mx1-cre Srsf2*fl/fl mice (Figure 1G; Figure S1I). Myeloid dysplasia was apparent based on detection of hypolobated and hypogranulated neutrophils, whereas erythroid dysplasia was evident based on nuclear irregularities and cytoplasmic vacuolization and blebbing in erythroid precursors. Overall, these results indicate that mutations in *Srsf2*P95H result in morphologic dysplasia and cytopenias with preserved marrow cellularity, features that are characteristic of human MDS, whereas complete loss of Srsf2 is incompatible with hematopoiesis.

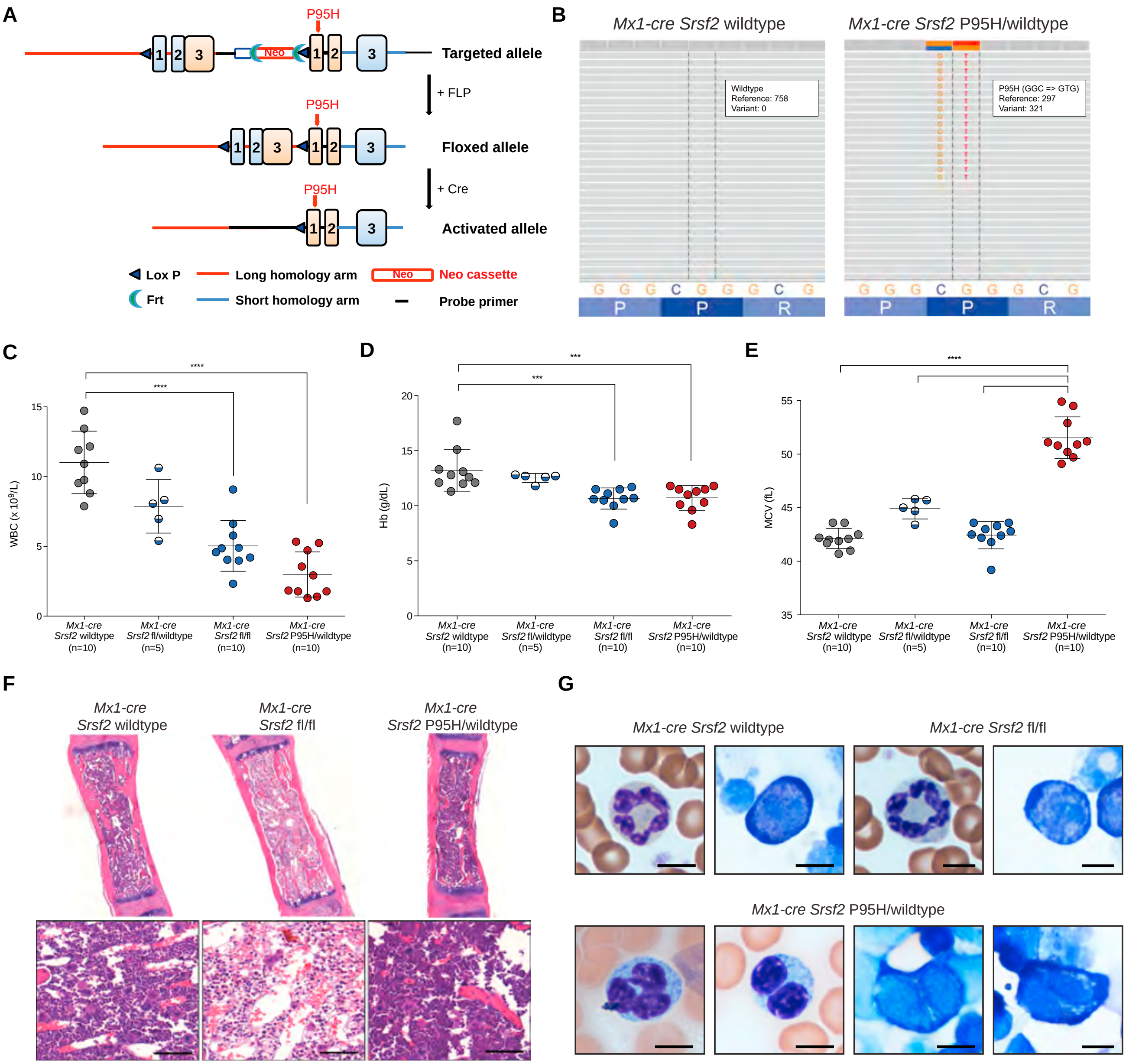


Figure 1. Conditional Expression of *Srsf2*P95H Results in Myeloid Dysplasia, a Phenotype Distinct from Heterozygous or Homozygous Loss of *Srsf2*

(A) Depiction of the *Srsf2*P95H allele.

(B) RNA-seq of LSK cells in *Mx1-cre Srsf2*WT and *Mx1-cre Srsf2*P95H/WT mice.

(C–E) White blood cell (WBC) count (C), hemoglobin (Hb) (D), and MCV (E) of red blood cells of CD45.1 recipient mice 18 weeks following noncompetitive transplantation of bone marrow from CD45.2+ *Mx1-cre Srsf2*WT, *Mx1-cre Srsf2*fl/WT, *Mx1-cre Srsf2*fl/fl, and *Mx1-cre Srsf2*P95H/WT mice (n = 10 mice/genotype for all genotypes except *Mx1-cre Srsf2*fl/WT, where n = 5; pIpC was administered to recipient mice 4 weeks following transplantation).

(F and G) H&E staining of femurs (scale bars, 50 μm) (F) and peripheral blood smears (G) from *Mx1-cre Srsf2*WT, *Mx1-cre Srsf2*fl/fl, or *Mx1-cre Srsf2*P95H/WT mice (scale bars, 10 μm). A representative neutrophil (left) and erythroid precursor (right) is shown for *Srsf2* WT and KO mice. *Mx1*-cre *Srsf2*P95H cells were marked by hypolobated and hypogranulated neutrophils (left two photos) and nuclear irregularities as well as cytoplasmic vacuolization and blebbing of erythroid precursors (right two photos).

Error bars represent mean ± SD. ***p < 0.001; ****p < 0.0001. See also Figure S1.

Given that mutations in *SRSF2* occur as early genetic events in MDS pathogenesis (Papaemmanuil et al., 2013) and that MDS is characterized by expansion of HSPCs, we next examined HSPC numbers and function in *Srsf2*P95H mice. Analysis of CD45.2+ HSPC subsets from *Mx1-cre Srsf2*P95H/WT mice and littermate controls 14 weeks after pIpC injection revealed expansion of lineage-negative Sca1+ c-Kit+ (LSK) and restricted hematopoietic progenitor cells (LSK CD48+ CD150+; hematopoietic

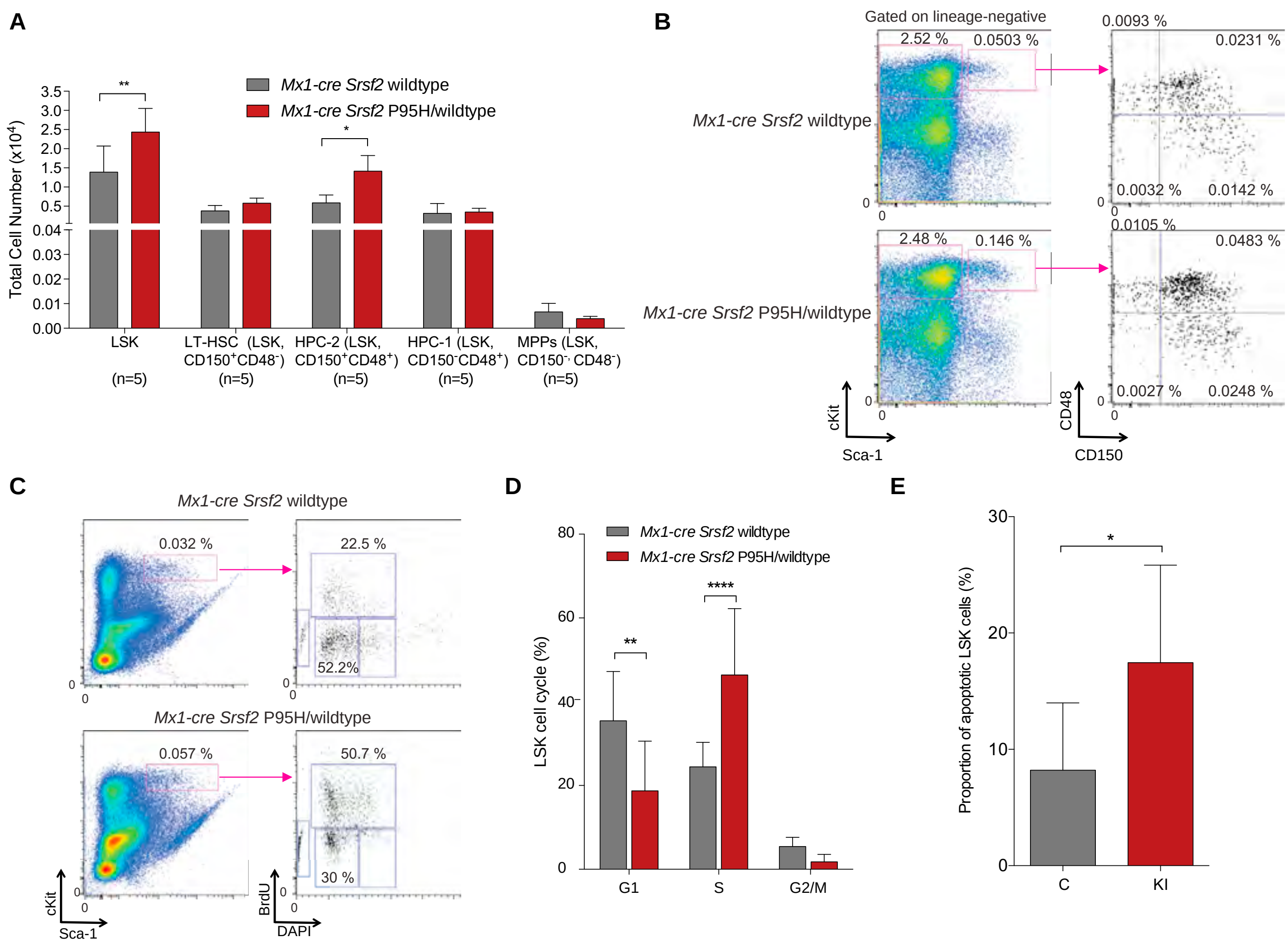


Figure 2. Conditional Expression of *Srsf2*P95H Results in Expansion of Hematopoietic Stem and Progenitor Cells with Increased Cell Proliferation and Apoptosis

(A and B) Enumeration (A) and fluorescence-activated cell sorting (FACS) analysis (B) of BM LSK cells, long-term hematopoietic stem cells (LT-HSC), restricted hematopoietic progenitor cell fractions 1 (HPC-1) and 2 (HPC-2), and multipotent progenitor (MPP) cells (Oguro et al., 2013) in 12-week-old *Mx1-cre Srsf2* WT and *Mx1-cre Srsf2*P95H/WT mice (n = 5 mice/genotype).

(C) Cell cycle analysis of LSK cells from *Mx1-cre Srsf2*WT or *Mx1-cre Srsf2*P95H/WT mice with in vivo bromodeoxyuridine (BrdU) administration. A representative FACS plot analysis shows gating on LSK cells followed by BrdU versus 4',6-diamidino-2-phenylindole (DAPI) stain (left).

(D) Relative quantification of the percentage of LSK cells in S, G2M, and G1 phase is shown on the right (n = 8 mice per group).

(E) Relative quantification of the percentage of Annexin V+/DAPI− LSK cells (n = 8 mice/genotype). C, control; KI, knockin.

Error bars represent mean ± SD. *p < 0.05, **p < 0.01, ****p < 0.0001. See also Figure S2.

progenitor cell fraction 2 [HPC-2]; Oguro et al., 2013) in mutant mice relative to controls (Figures 2A and 2B). A similar LSK expansion was seen in spleens of *Srsf2*P95H mutant mice (although splenomegaly was not observed up to 20 weeks post-pIpC) (Figures S2A and S2B). Because the detection of increased HSPCs in *Srsf2*P95H mutant mice appeared paradoxical given the decreased peripheral blood counts in these same mice, we next examined the cell cycle kinetics and apoptosis of *Srsf2* mutant HSPCs. Indeed, *Srsf2*P95H LSK cells were characterized by an increase in the proportion of cells in S-phase as well as in early apoptosis (Figures 2C–2E). Despite HSPC expansion in *Srsf2*P95H mutant mice, purified LSK cells from mice with a homozygous *Srsf2* deletion or heterozygous *Srsf2*P95H mutation had similarly impaired colony formation and serial re-plating capacity in vitro (Figure S2C).

To assess the functional effects of *Srsf2* alterations on HSC self-renewal in vivo, we next compared *Srsf2* heterozygous KO, homozygous KO, and heterozygous P95H mutant mice in competitive transplantation assays (Figure 3A). Equal numbers of BM MNCs from CD45.1 WT mice and CD45.2 *Mx1-cre Srsf2* WT, *Mx1-cre Srsf2*fl/WT, *Mx1-cre Srsf2*fl/fl, or *Mx1-cre Srsf2*P95H/WT mice were transplanted into lethally irradiated CD45.1 mice, followed by pIpC injection 4 weeks later. An assessment of peripheral blood chimerism monthly thereafter revealed a complete loss of CD45.2 chimerism in mice transplanted with *Mx1-cre Srsf2*fl/fl cells and a significant decrease in chimerism in mice transplanted with *Mx1-cre Srsf2*P95H/WT cells (Figure 3B; Figures S3A and S3B). However, an analysis of BM LSK chimerism 18 weeks post-transplant revealed an increase in CD45.2+ HSPCs derived from *Srsf2*P95H mice relative to other

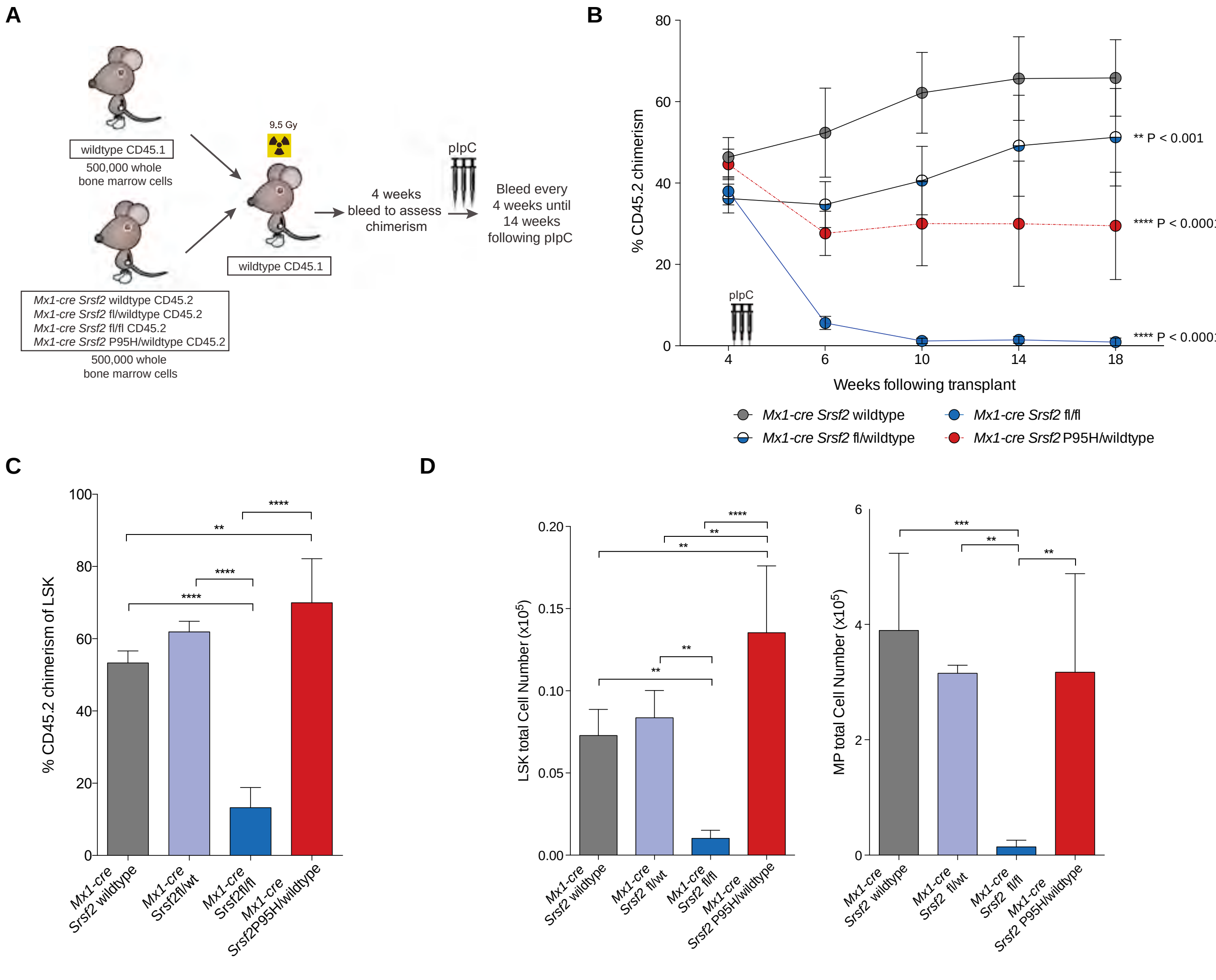


Figure 3. *Srsf2*P95H Mutation Impairs Hematopoietic Stem Cell Self-Renewal in a Manner Distinct from *Srsf2* Loss

(A) Depiction of a competitive BM transplantation assay. pIpC, polyinosinic-polycytidylic acid.

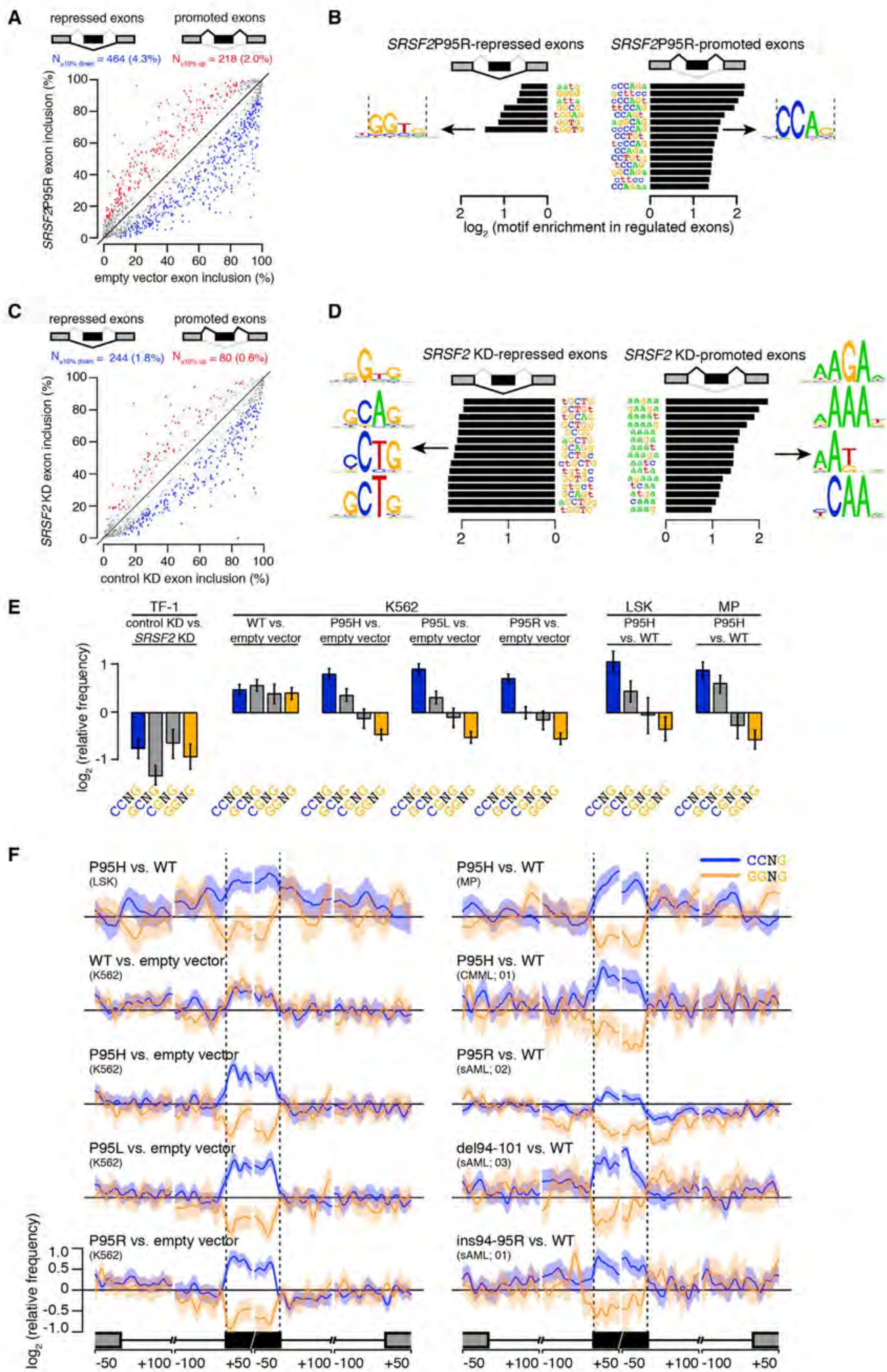
(B) Percentage of CD45.2+ chimerism in the peripheral blood of recipient mice (n = 10 mice/genotype).

(C and D) Chimerism (C) and flow cytometric enumeration (D) of CD45.2+ LSK (left) and MP (lineage-negative Sca1-c-Kit+, right) cells in the BM of *Mx1-cre Srsf2*WT, *Mx1-cre Srsf2*fl/WT, *Mx1-cre Srsf2*fl/fl, and *Mx1-cre Srsf2*P95H/WT mice 14 weeks after pIpC injection.

Error bars represent mean ± SD. **p < 0.001, ***p < 0.0002, ****p < 0.0001. See also Figure S3.

groups and a near complete absence of CD45.2+ HSPCs from *Srsf2*fl/fl mice (Figures 3C and 3D; Figure S3C). Serial competitive transplantation of whole bone marrow from *Srsf2*P95H, *Srsf2* heterozygous KO, and *Srsf2* WT primary recipient transplanted mice continued to reveal an impaired reconstitution capacity of *Srsf2*P95H mutant mice relative to *Srsf2* heterozygous KO or control mice (Figure S3D). Of note, colony assays and competitive transplantation experiments were performed using multiple genotypes of control mice (*Cre*-negative *Srsf2*P95H mice as well as *Mx1*-cre *Srsf2* WT mice; Figures S2C and S3E) to control for any possible confounding effect of Cre expression or the presence of the unexcised P95H knockin allele.

The fact that *Mx1-cre Srsf2*P95H/WT mice had an increase in HSPCs despite impaired formation of mature peripheral blood cells suggested that mutant *Srsf2* was associated with impaired HSPC differentiation. Flow cytometric analysis of mature and intermediate precursor cell subsets in *Srsf2*P95H mice was therefore performed to identify the stage of impaired hematopoiesis. This revealed that peripheral leukopenia was predominantly due to decreased peripheral blood B cells, evident at all stages of B lymphopoiesis following the transition of pre-proB to proB cells, in *Srsf2*P95H mice relative to controls (Figures S3F and S3G). Moreover, immunophenotypic analysis of intermediate hematopoietic progenitors (Pronk et al., 2007) revealed deficits in early erythroid progenitors in *Srsf2*P95H mice relative to controls, initiating at the pre-MegE and pre-colony-forming units, erythroid, stages (Figures S3H and S3I). Given prior data showing that homozygous deletion of *Srsf2* resulted in defective T cell maturation and CD45 splicing (Wang et al., 2001), we also examined thymic T cell differentiation and CD45 isoform expression in *Srsf2*P95H mice relative to controls (Figures S3J and S3K). This revealed no effect of *Srsf2*P95H mutation on thymic



(legend on next page)

T cell maturation or protein expression of the specific CD45 isoforms identified previously to be downregulated with homozygous deletion of *Srsf2* (Wang et al., 2001).

Collectively, the biological analysis of *Srsf2*P95H mutant mice identified phenotypes distinct from mice with a partial or complete loss of *Srsf2*, suggesting that *SRSF2* mutations alter SRSF2's normal function rather than resulting in haploinsufficiency or a dominant-negative function. Of note, despite the impaired hematopoietic differentiation, increase in HSPC subsets, and morphologic dysplasia in *Srsf2*P95H/WT mice, no *Srsf2*P95H mutant mice developed acute myeloid leukemia in up to 70 weeks of observation.

SRSF2 Mutations Are Associated with Global Alterations of Gene Expression and Splicing

We next sought to identify the transcriptional and post-transcriptional alterations caused by *SRSF2* mutations through RNA-seq of purified LSK and myeloid progenitor (MP, lineage-negative Sca1- c-Kit+) populations. This was performed 4 weeks after pIpC administration. In an unsupervised cluster analysis based on coding gene expression, samples clustered first by cell type and then by genotype (Figure S4A). The expression of several hematopoietic regulators was altered in *Srsf2*P95H mutant cells, including upregulation of *Gfi1*, *Cebpe*, and *Hoxb2* in LSK cells; downregulation of *Gata1* and *Gata2* in MP cells; and downregulation of *Cdkn1a* in both populations. In addition, we observed preferential down- versus upregulation of the expression of coding genes in *Srsf2* mutant cells relative to the WT (Figures S4B and S4C). Gene ontology (GO) analysis revealed an enrichment for the downregulation of genes in both LSK and MP cells involved in the regulation of cell cycle, proliferation, differentiation, and apoptosis (upregulated genes were not enriched for these processes; Figure S4D).

To identify changes in splicing driven by *SRSF2* mutations that might contribute to disease, we augmented our mouse data with RNA-seq data from primary CMML (n = 13; 3 with *SRSF2* mutation) and AML (n = 9, 5 with *SRSF2* mutation) patient samples (Table S1) as well as K562 cells ectopically expressing an empty vector or a single allele of *SRSF2* (WT, P95H, P95L, and P95R). In all sequenced patients with *SRSF2* mutations, the WT and mutant alleles were expressed at similar levels (Table S1), as was the case for the *Srsf2*P95H mouse cells (Figure 1). Similarly, isogenic K562 cells with lentiviral expression of WT or mutant SRSF2 cells expressed WT and mutant SRSF2 at roughly equal levels (Figures S4E–S4G). We quantified global changes in splicing of ~125,000 alternative splicing events and ~160,000 constitutive splice junctions associated with *SRSF2* mutations in these five datasets (LSK, MP, CMML, AML, and K562). We required a minimum change in isoform ratio of 10% to call an event differentially spliced (where a change in isoform ratio is defined as an absolute, rather than relative, quantity as the increase or decrease in the percentage of all mRNAs transcribed from the parent gene that follow a given splicing pattern). In all datasets, *SRSF2* mutations were associated with differential splicing of all classes of splicing events as well as novel alternative splicing and intron retention of splice junctions annotated as constitutively spliced. However, only a relatively small fraction of alternatively spliced events of any class were affected by *SRSF2* mutations (Figure S4H). *SRSF2* mutations were associated with a mild bias toward exon skipping but did not lead to globally increased levels of predicted substrates for degradation by nonsense-mediated decay.

SRSF2 Mutations Alter Exonic Splicing Enhancer Preference but SRSF2 Loss Does Not

Because SRSF2 normally recognizes ESE elements within the pre-mRNA to promote exon recognition (Graveley and Maniatis, 1998; Liu et al., 2000; Schaal and Maniatis, 1999; Zahler et al., 2004), we hypothesized that *SRSF2* mutations might alter its normal sequence-specific activity. To test this, we performed an ab initio motif identification screen. We quantified the occurrence of each possible k-mer (k = 4, 5, 6) within cassette exons that were differentially spliced in *Srsf2*P95H MP cells and identified k-mers that were enriched or depleted in cassette exons promoted versus repressed in *Srsf2*P95H cells. We identified enriched and depleted motifs using a non-parametric (Kolmogorov-Smirnov) statistical test with a p value threshold of 0.05. Significantly enriched k-mers were C-rich, whereas depleted k-mers were G-rich (Figures S4I and S4J). We then performed an identical analysis using our K562 data, which likewise identified CCAG and GGTG as the most enriched and depleted consensus motifs, respectively (Figures 4A and 4B). A recent solution structure of SRSF2 in complex with RNA revealed that SRSF2 has a consensus motif of SSNG (where "S" represents

Figure 4. *SRSF2* Mutations Alter Exonic Splicing Enhancer Preference

(A) Scatterplot of cassette exon inclusion in K562 cells expressing empty vector or *SRSF2*P95R. Percentages indicate the percent of alternatively spliced cassette exons with increased or decreased inclusion. Red and blue dots represent individual cassette exons that are promoted or repressed in *SRSF2*P95R versus empty vector cells, respectively. Promoted and repressed cassette exons are defined as those whose inclusion levels are increased or decreased by $\geq$10% with a Bayes factor of $\geq$5, as estimated by Wagenmakers' framework (Wagenmakers et al., 2010).

(B) Enriched (right) and depleted (left) k-mers in cassette exons promoted versus repressed in *SRSF2*P95R versus WT cells.

(C) Scatterplot of cassette exon inclusion in TF-1 cells following transfection with a siRNA against *SRSF2* or a control non-targeting siRNA (KD, knockdown). Percentages indicate the percent of alternatively spliced cassette exons with increased or decreased inclusion.

(D) Enriched (right) and depleted (left) k-mers in cassette exons promoted versus repressed in *SRSF2* KD versus control cells.

(E) Mean enrichment of all variants of the SSNG motif in cassette exons promoted versus repressed in TF-1 cells following *SRSF2* knockdown and K562, LSK, and MP cells expressing WT or mutant *SRSF2*. Error bars indicate 95% confidence intervals estimated by bootstrapping.

(F) Relative frequency of CCNG and GGNG motifs in cassette exons promoted versus repressed by *SRSF2* mutations in LSK and MP cells (top), K562 cells (left), and primary AML and CMML samples with or without *SRSF2* mutations (right) (the sample numbers correspond to the patient identifiers in Table S1). Shading indicates 95% confidence interval by bootstrapping. The schematic illustrates a portion of a metagene containing the differentially spliced cassette exon. From left to right, the features are the upstream exon (gray box) and intron (black line), the cassette exon (black box, vertical dashed lines), and the downstream intron (black line) and exon (gray box). Horizontal axis, genomic coordinates defined with respect to the 5′ and 3′ splice sites where 0 is the splice site itself. Vertical axis, relative frequency of the indicated motifs over genomic loci containing cassette exons promoted versus repressed by *SRSF2* mutations (log scale).

See also Figure S4.

C or G) and efficiently recognizes both CCNG and GGNG (Daubner et al., 2012). Therefore, our ab initio analysis suggested that mutations affecting the P95 residue may alter SRSF2's ability to recognize variants of its normal SSNG motif.

To further explore this hypothesis, we compared the relative enrichment of all four SSNG variants in cassette exons that were differentially spliced upon depletion of *SRSF2*, overexpression of WT *SRSF2*, or expression of mutant *SRSF2*. *SRSF2* depletion—achieved by knockdown of endogenous *SRSF2* in the absence of mutant protein expression (Figure S4K)—caused preferential skipping of cassette exons, consistent with SRSF2's canonical role in promoting exon recognition (Figure 4C). Ab initio motif analyses identified both C- and G-rich variants of the SSNG motif as the most enriched motifs in cassette exons that were repressed following SRSF2 depletion (Figure 4D). Quantitation of the enrichment of each SSNG variant revealed that all were associated with exon repression following knockdown. In contrast, overexpression of WT SRSF2 was associated with enrichment of each SSNG variant (Figure 4E). These data suggest that different SSNG variants function as equally efficacious SRSF2-dependent ESEs, consistent with SRSF2's in vitro binding specificity (Daubner et al., 2012). In contrast, K562 cells as well as LSK and MP cells expressing mutant *Srsf2* exhibited enrichment for CCNG and depletion for GGNG in exons that were promoted versus repressed (Figure 4E).

To test whether this motif enrichment and depletion was due to ESE activity, we computed the spatial distribution of CCNG and GGNG motifs across genomic loci containing cassette exons that were promoted or repressed in association with *SRSF2* mutations. CCNG and GGNG were, respectively, enriched and depleted specifically over cassette exons and not over the flanking introns or exons. We observed similar motif preferences and distributions in patient transcriptomes (Figure 4F). Because CCNG/GGNG motifs were not consistently enriched/depleted in introns flanking differentially spliced cassette exons, and because we were unable to identify enriched motifs with *ab initio* searches in introns, we conclude that differential cassette exon splicing is likely due primarily to altered recognition of exonic motifs. Together, these data reveal spatially restricted enrichment of specific ESEs in association with *SRSF2* mutations and suggest that *SRSF2* mutations cause alteration rather than loss of normal ESE recognition activity.

SRSF2 Proline 95 Mutations Alter RNA Binding Specificity by Changing the Conformation of Both RRM Termini

We next tested whether this association between *SRSF2* mutations and enrichment/depletion of specific ESEs was due to altered SRSF2:RNA interactions. We purified SRSF2's RNA RRM as described previously and performed isothermal titration calorimetry (ITC) with the RNA ligand 5′-uCCAGu-3′, an optimal SRSF2 target according to the SSNG consensus sequence (Daubner et al., 2012). All three P95 mutations resulted in an increase in binding affinity of 3.9- to 4.5-fold relative to WT SRSF2 (Figures 5A and 5B; Figure S5A), consistent with the enrichment for CCNG motifs that we observed in exons promoted by *SRSF2* mutations (Figure 4B). We next tested whether P95 mutations resulted in altered RNA binding specificity. In contrast to 5′-uCCAGu-3′ RNA, ITC measurements revealed that all three P95 mutants exhibited a 1.2- to 2.1-fold decrease in binding affinity to the 5′-uGGAGu-3′ RNA relative to WT SRSF2 (Figures 5A and 5B; Figure S5B). ITC measurements using the RNA sequences 5′-uGCAGu-3′ and 5′-uCGAGu-3′ revealed that G > C substitutions at the second motif position resulted in larger increases in binding affinity than at the first motif position (2.6- to 3.4-fold versus 1.1- to 1.8-fold; Figure 5B; Figures S5C and S5D). The RNA binding preferences measured by ITC were remarkably consistent with the ESE enrichment identified by RNA-seq. For each mutant, the level of motif enrichment (Figure 4E) was roughly proportional to the affinity increase (Figure 5C), and the enrichment and affinity measurement supported the same relative preference for each specific motif (CC > GC > CG > GG). This strongly supports the notion that the splicing changes caused by P95 mutations are the result of an altered sequence specificity of the SRSF2 RRM.

P95 is located at the C-terminal end of the SRSF2 RRM, and the published solution structure of SRSF2 in complex with 5′-uCCAGu-3′ revealed extensive contacts of P95 with the second cytosine (Figure S5E), emphasized by several intermolecular nuclear Overhauser effects (NOEs) (Daubner et al., 2012). To test whether SRSF2's RNA binding surface was altered by P95 mutations, we conducted nuclear magnetic resonance (NMR) titration with the SRSF2 P95H RRM and the 5′-uCCAGu-3′ RNA and assigned the backbone of this complex using standard heteronuclear NMR experiments. Mapping of the chemical shift perturbations revealed that the RNA-binding surface of the RRM is not disturbed by the P95H mutation. However, both termini experienced large changes in their environment (Figure 5D), an observation that held true for all three P95 mutations (Figure S5F). Consistent with our ESE and ITC analyses, this relocation of termini primarily affected the second cytosine, which exhibited the largest chemical shift perturbations of its proton resonances (Figures S5G and S5H). Smaller changes of chemical shifts were observed when P95 mutants were bound to 5′-uGGAGu-3′ (Figure S5H). Together, our experiments indicate that *SRSF2* mutations change SRSF2's normal RNA-binding affinity and specificity in vitro, likely explaining the widespread alterations in ESE preference we observed in vivo.

Mutant *SRSF2* Promotes Mis-splicing and Degradation of *EZH2*

We next used our transcriptome data to identify common changes in splicing driven by *SRSF2* mutations that might contribute to disease. Intersection of differentially spliced genes in LSK, MP, CMML, and AML samples identified 75 genes differentially spliced in association with *SRSF2* mutations in both LSK and MP cells and at least one primary patient cohort as well as an additional 97 (LSK) and 87 (MP) genes differentially spliced in one mouse cell population, but not the other, as well as a patient cohort (Figure 6A; Tables S2–S5). Many of these genes have a known importance in myeloid malignancies. For example, *SRSF2* mutations promoted the inclusion of a highly conserved "poison" cassette exon of *EZH2* (*Enhancer of zeste homolog 2*) and repressed a frame-preserving cassette exon of *BCOR* (*BCL6 corepressor*) (Figure 6A; Figures S6A and S6B). Of note, we did not identify altered splicing of CD45 in *SRSF2* mutant cells (Tables S2–S5), which has been noted previously as being

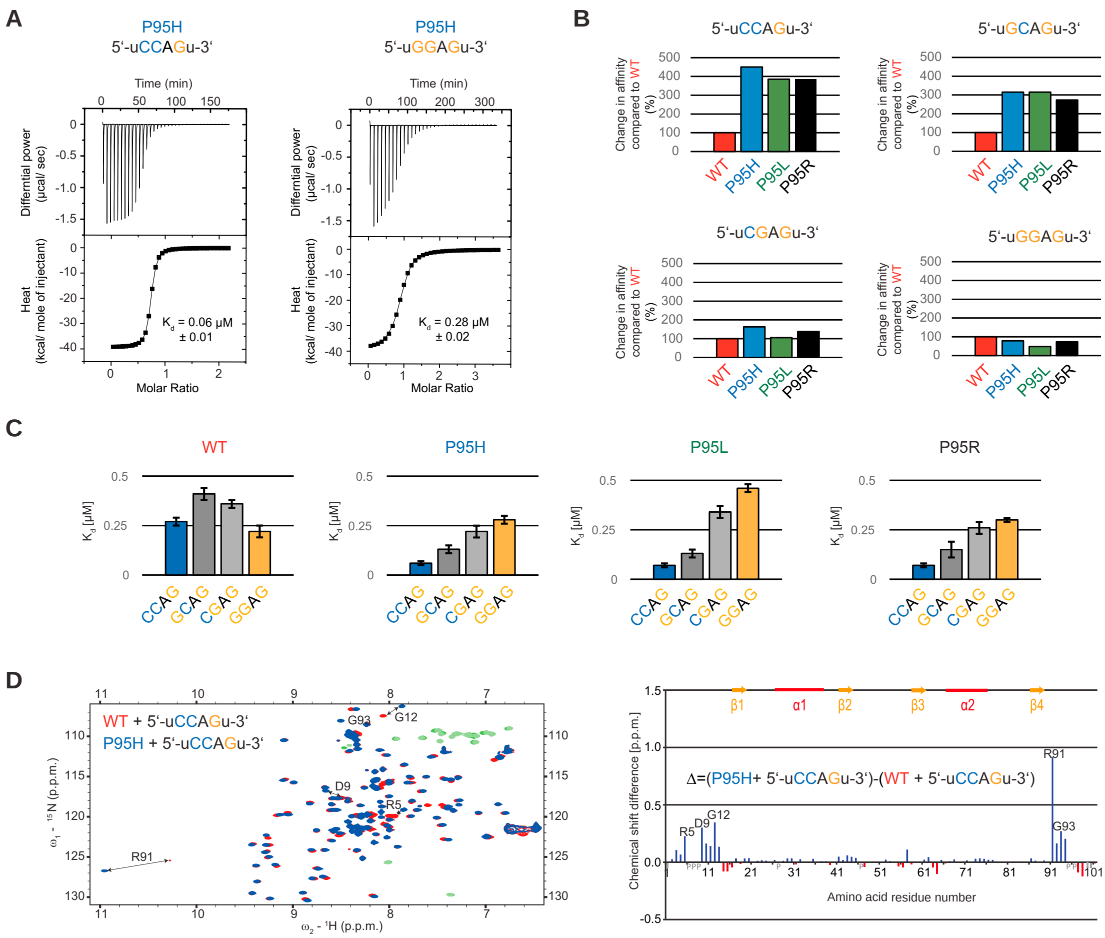


Figure 5. Proline 95 Mutations Change the RNA-Binding Specificity of the SRSF2 RNA RRM In Vitro and Lead to Relocation of the N and C Termini

(A) ITC raw data and binding curve for the SRSF2 RRM P95H mutant with 5′-uCCAGu-3′ and 5′-uGGAGu-3′ RNA.

(B) Change in RNA-binding affinity (percent) for SRSF2 RRM P95H (blue), P95L (green), and P95R (black) mutants compared with WT (red) (Daubner et al., 2012) using RNA targets 5′-uCCAGu-3′, 5′-uGCAGu-3′, 5′-uCGAGu-3′, and 5′-uGGAGu-3′.

(C) Change in RNA-binding specificity of SRSF2 RRM WT, P95H, P95L, and P95R with 5′-UCCAGU-3′ (blue), 5′-UGCAGU-3′ (dark gray), 5′-UCGAGU-3′ (light gray), and 5′-UGGAGU-3′ RNA (orange). Error bars represent mean ± SD.

(D) Left: overlay of 2D [^{15}N-^{1}H] heteronuclear single quantum coherence (HSQC) of the wild-type (red) and P95H mutant (blue) bound to 5′-UCCAGU-3′ RNA, with negative peaks shown in green (WT) and light green (mutant). Right: difference of the chemical shift perturbations of the P95H mutant and wild-type. Positive values (blue) with a higher perturbation with the P95H mutant and negative values (red) with a higher perturbation with the WT are shown. Missing assignments are marked with gray bars and proline with a gray P. Residues with the highest difference are depicted in both the graph and spectra.

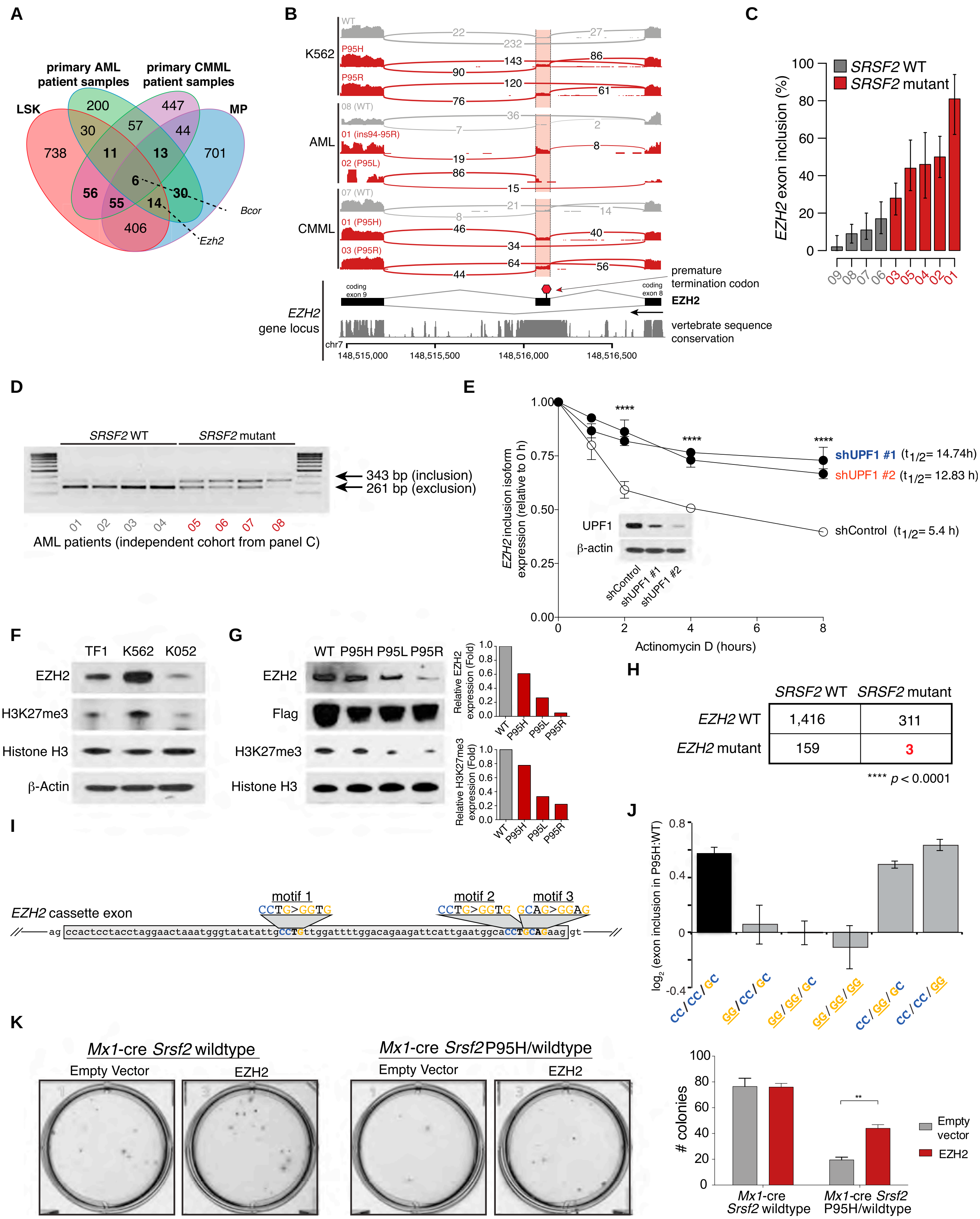
See also Figure S5.

altered in murine *Srsf2* KO hematopoietic cells (Wang et al., 2001).

To identify potential functional consequences of recurrent mis-splicing, we focused on the splicing event in *EZH2*. *SRSF2* mutant cells exhibited preferential inclusion of a poison cassette exon that introduces a premature termination codon predicted to result in nonsense-mediated decay (NMD) of *EZH2* (Figures 6B and 6C). Both the poison exon itself and its flanking intronic sequences exhibited high sequence conservation across vertebrates, exceeding the sequence conservation exhibited by the upstream and downstream constitutive coding exons themselves, which is a common feature of physiologically important splicing events (Lareau et al., 2007; Ni et al., 2007; Figure 6B).

We validated this *EZH2* splicing change using both qualitative and quantitative isoform-specific RT-PCR in leukemia cell lines that were WT or mutant for *SRSF2* (Figures S6C and S6D) as well as in an independent panel of primary AML patient samples with or without *SRSF2* mutations (n = 8, 4 with *SRSF2* mutations; Figure 6D; Figure S6E).

Next, to confirm whether the cassette exon promoted by *SRSF2* mutations triggers degradation by NMD, we measured the half-life of the inclusion isoform of *EZH2* in *SRSF2*P95H cells transfected with a control or anti-*UPF1* (a required NMD factor) short hairpin RNA (shRNA) following transcriptional shutoff with actinomycin D ('t Hoen et al., 2011; Figure 6E; Figures S6F and S6G). The fact that the mRNA half-life of the inclusion isoform of *EZH2* was lengthened by *UPF1* knockdown in these



(legend on next page)

experiments suggests that this particular isoform of *EZH2*, which is promoted by mutant SRSF2, undergoes NMD. The half-life of a well-characterized NMD substrate of *SRSF3* (Lareau et al., 2007; Ni et al., 2007) increased similarly following *UPF1* knockdown, confirming that *UPF1* knockdown effectively inhibited NMD (Figure S6H).

Next, to identify whether the protein product of *EZH2* is altered in *SRSF2* mutant cells, we performed WB analysis of a panel of human AML cell lines WT (TF-1, K562) or mutant for *SRSF2* (K052) (all WT for *EZH2*). This revealed lower EZH2 protein levels as well as lower global levels of histone H3 lysine 27 trimethylation (H3K27me3, a methylation mark placed by EZH2) in *SRSF2* mutant K052 cells (Figure 6F). To further validate this finding in an isogenic context, we performed WB analysis in K562 cells ectopically expressing WT *SRSF2* or *SRSF2*P95H/L/R mutant cDNA. This analysis revealed a consistent downregulation of EZH2 protein expression as well as global H3K27me3 in all three *SRSF2* mutant samples compared with *SRSF2* WT K562 cells (Figure 6G).

Consistent with *SRSF2* mutations promoting a disabling splicing change in *EZH2*, *EZH2* loss-of-function mutations are common in MDS. In an analysis of >1,800 MDS patients where *EZH2* and *SRSF2* were both sequenced, *EZH2* loss-of-function mutations were mutually exclusive with *SRSF2* mutations ($p < 0.0001$) (Bejar et al., 2012; Ernst et al., 2010; Haferlach et al., 2014; Muto et al., 2013; Papaemmanuil et al., 2013; Figure 6H).

The above data strongly link *SRSF2* mutations to disabling splicing of *EZH2*. We next sought to examine whether the change in RNA ESE preference induced by *SRSF2* mutations caused *EZH2* mis-splicing. We therefore cloned the genomic locus containing the *EZH2* poison exon and flanking introns and constitutive exons to create a minigene that recapitulates this splicing event. We identified three potential SRSF2-dependent SSNG motifs in the poison exon (CCTG, CCTG, and GCAG), one or more of which we expected to be better recognized by mutant SRSF2 than WT SRSF2. We then mutated each motif to the corresponding GG equivalent, both separately and in combination (Figure 6I). Measuring cassette exon recognition in K562 cells expressing WT or mutant *SRSF2*, we found that the first motif was required for robust splicing change in *SRSF2* mutant cells, such that the mutation CCTG > GGTG prevented an increase in poison exon recognition (Figure 6J). We conclude that *SRSF2* mutations induce a disabling splicing change in *EZH2* in an ESE-dependent manner, consistent with altered RNA recognition activity.

We next sought to test whether restoring normally spliced *EZH2* mRNA could rescue hematopoiesis in *SRSF2* mutant cells. *EZH2* full-length cDNA or an empty vector (both in a retroviral ZsGreen1 vector) were overexpressed in c-Kit+ *Srsf2*P95H or WT cells, followed by assessment of methylcellulose colony formation of c-Kit+/ZsGreen1+ cells. *EZH2* cDNA was equally overexpressed in *Srsf2* mutant and WT cells (Figure S6I), and *Srsf2*P95H mutant cells overexpressing full-length *EZH2* experienced an ~50% increase in colony formation relative to *Srsf2*P95H mutant cells expressing an empty vector (Figure 6K; Figure S6J). In contrast, *EZH2* overexpression had no substantial effect on initial colony formation in *Srsf2* WT cells (Figure 6K; Figure S6I). These data identify that restoration of normally spliced *EZH2* mRNA in *SRSF2* mutant cells at least partially rescues the hematopoietic defects induced by mutant *SRSF2.*

DISCUSSION

The consistent occurrence of heterozygous point mutations affecting highly restricted residues of spliceosomal proteins strongly suggests a gain-of-function or dominant-negative activity for these mutations in malignant transformation. Here we identify an effect of the *SRSF2*P95H mutation distinct from loss of *SRSF2* and reveal that mutations in *SRSF2* confer an alteration in function that results in key aspects of MDS. This includes an increase in HSPCs in *Srsf2*P95H mutant mice with impaired differentiation, altered cell cycle kinetics, and increased apoptosis resulting in peripheral cytopenias and morphologic dysplasia. By contrast, WT *Srsf2* appears to be constitutively required for hematopoiesis.

Figure 6. *SRSF2* Mutant Primary Murine and Patient Samples Exhibit Convergent Splicing Alterations

(A) Intersection of genes exhibiting differential splicing in *SRSF2* mutant versus WT mouse LSK and MP cells and primary AML and CMML samples (restricted to orthologous genes).

(B) Integrative Genomics Viewer (IGV)/Sashimi plot illustrating the *EZH2* cassette exon promoted by *SRSF2* mutations in multiple datasets analyzed here (top) (the patient numbers listed in the Sashimi plot correspond to the numbers in Table S1 detailing patient characteristics). The DNA sequence conservation of the locus, as estimated by phastCons (Siepel et al., 2005), across 30 vertebrate species is shown in the track below the Sashimi plot.

(C) Bar plot describing the percentage of *EZH2* transcripts harboring a specific cassette exon in the *SRSF2* mutant relative to WT primary AML samples from RNA-seq data. Error bars indicate 95% confidence intervals.

(D) RT-PCR of an *EZH2* exon inclusion event in an independent set of *SRSF2* WT and mutant AML samples.

(E) Quantitative RT-PCR of *EZH2* inclusion isoform in *SRSF2*P95H mutant cell line K052 cells with or without UPF1 knockdown and actinomycin D treatment.

(F) WB analysis for EZH2 and H3K27me3 in *SRSF2/EZH2* WT (TF-1, K562) and *SRSF2*P95H mutant/*EZH2* WT (K052) AML cell lines.

(G) WB analysis for EZH2, H3K27me3, and FLAG epitope in K562 cells with lentiviral overexpression of N-terminal FLAG-tagged SRSF2 WT, SRSF2P95H, SRSF2P95L, or SRSF2P95R (left). Relative quantification of EZH2 protein expression by WB to total histone H3 expression in K562 cells expressing SRSF2 mutants relative to WT is shown on the right.

(H) *EZH2* and *SRSF2* mutations are mutually exclusive in the sequencing of DNA from >1,000 MDS patients (Bejar et al., 2012; Ernst et al., 2010; Haferlach et al., 2014; Muto et al., 2013; Papaemmanuil et al., 2013).

(I) Schematic of the *EZH2* cassette exon with SSNG motifs highlighted and mutations to GG equivalents shown.

(J) *EZH2* cassette exon inclusion for minigenes containing the endogenous cassette exon or a cassette exon with mutation of motifs 1, 2, and/or 3 to the GG equivalent.

(K) Photographs (left) and enumeration (right) of c-Kit+/ZsGreen1+ cells from *Srsf2* WT or *Srsf2*P95H mice 14 days after overexpression of empty vector or *EZH2* cDNA and plating in methylcellulose medium.

$^{**}p < 0.01$, $^{****}p < 0.0001$. Error bars represent mean ± SD unless stated otherwise. See also Figure S6 and Tables S1–S5.

Transcriptional analysis of *SRSF2* mutant cells revealed that *SRSF2* mutations result in genome-wide alterations in ESE preference in both human and murine cells. Biochemical analysis of the interaction of SRSF2 with RNA in cell-free in vitro assays identified an analogous change in specificity of interactions between SRSF2 and pre-mRNA induced by *SRSF2* mutations. This altered interaction of mutant SRSF2 with RNA appears to be due to an effect of *SRSF2*P95H/L/R mutations on the conformations of the termini of SRSF2's RRM domain, as revealed by NMR spectroscopy. Our genomic and biochemical assays indicate that *SRSF2* mutations cause alteration rather than loss-of-function, driving preferential recognition of cassette exons containing C- versus G-rich ESEs.

The altered pre-mRNA recognition activity of mutant SRSF2 likely underlies the mis-splicing of key transcriptional regulators—several of which have been implicated previously in MDS pathogenesis. This includes promotion of a poison exon of *EZH2* that undergoes NMD and results in reduced EZH2 protein expression in *SRSF2* mutant cells. Loss-of-function mutations in *EZH2* occur in the same exact spectrum of myeloid malignancies as *SRSF2* mutations (Ernst et al., 2010; Nikoloski et al., 2010) and loss of *Ezh2* has been functionally linked to MDS development in vivo (Muto et al., 2013). Moreover, *SRSF2* and *EZH2* mutations are mutually exclusive in MDS patients (Haferlach et al., 2014; Papaemmanuil et al., 2013), but the basis for this observation was previously unknown. The data here provide a mechanistic basis for this mutual exclusivity as *SRSF2* mutations functionally reduce EZH2 protein expression.

In addition to the effects of mutant *SRSF2* on *EZH2* splicing and protein expression, a number of other genes of known importance in hematopoiesis and malignancy were also consistently differentially spliced in isogenic human cells, primary patient samples, and murine cells bearing mutant *SRSF2*. These include additional genes mutated in MDS (such as *BCOR*), genes with an importance in hematopoietic stem cell self-renewal (such as *IKAROS*), and genes critical for cell survival (such as *CASPASE 8*). Future efforts to understand the functional effects of each of these specific splicing events will be important in further delineating the effects of mutant *SRSF2* on MDS pathogenesis as well as possibly providing novel means for therapeutic targeting of *SRSF2* mutant cells.

Our studies, which reveal both mechanistic splicing alterations and specific mis-spliced isoforms in *SRSF2* mutant cells, may provide insights into therapeutic opportunities for targeting *SRSF2* mutant cells. For example, the observations that mutant *SRSF2* promotes the inclusion of a poison exon in an ESE-dependent manner and that restoration of normally spliced *EZH2* mRNA partially rescues defective hematopoiesis in *SRSF2* mutant cells suggest that normal cellular function may be at least partially restored by manipulating specific pathologic splicing events.

EXPERIMENTAL PROCEDURES

Generation of the *Srsf2*P95H conditional knockin mice is described in the Supplemental Experimental Procedures. All animal procedures were conducted in accordance with the Guidelines for the Care and Use of Laboratory Animals and approved by the Institutional Animal Care and Use Committees at Memorial Sloan Kettering Cancer Center.

Patient Samples

Studies were approved by the Institutional Review Boards of Memorial Sloan Kettering Cancer Center and Fred Hutchinson Cancer Research Center and conducted in accordance to the Declaration of Helsinki protocol. Informed consent was obtained from all human subjects.

mRNA Sequencing

For sorted mouse cell populations, K562 cells, and primary AML and CMML samples, RNA was extracted using QIAGEN RNeasy columns. poly(A)-selected, unstranded Illumina libraries were prepared with a modified TruSeq protocol. 0.5× AMPure XP beads were added to the sample library to select for fragments of <400 bp, followed by 1× beads to select for fragments of >100 bp. These fragments were then amplified with PCR (15 cycles) and separated by gel electrophoresis (2% agarose). 300-bp DNA fragments were isolated and sequenced on the Illumina HiSeq 2000 (100 million 2 × 49 bp reads/sample).

RNA-Seq Read Mapping

Reads were mapped to the University of California, Santa Cruz (UCSC) hg19 (NCBI GRCh37) human genome or UCSC mm10 (NCBI GRCm38) genome assemblies. First, a modified version of RNA-seq by expectation maximization (RSEM) that called Bowtie v1.0.0, with the −v 2 argument was created. This modified RSEM was then called with the arguments '--bowtie-m 100--bowtie-chunkmbs 500 --calc-ci --output-genome-bam' on the gene annotation file. Read alignments with MAPping Quality (MAPQ) scores of 0 and/or a splice junction overhang of less than 6 bp were then filtered out. The remaining unaligned reads were then aligned by TopHat v2.0.8b with the arguments '--bowtie1 --read-mismatches 2 --read-edit-dist 2--no-mixed --no-discordant --min-anchor-length 6 --splice-mismatches 0 --min-intron-length 10 --max-intron-length 1000000 --min-isoform-fraction 0.0 --no-novel-juncs --no-novel-indels --raw-juncs' on the splice junction file (--mate-inner-dist and --mate-std-dev were calculated by mapping to constitutive coding exons with the Mixture of Isoforms (MISO) exon_utils.py utility). The resulting TopHat alignments were then filtered as for the RSEM-generated alignments. Finally, the RSEM- and TopHat-created binary sequence alignment/map (BAM) files were merged to create final BAM files.

Isoform Expression Measurements

Two different methods were used to quantify isoform ratios. For alternative splicing events from MISO's v2.0 annotation, MISO was used to estimate isoform ratios. For alternative splicing or intron retention of annotated constitutive junctions, junction reads alone were used as described previously (Hubert et al., 2013). To identify differentially expressed events, we required a minimum of 20 identifying reads (supporting either, but not both, isoforms) per event as well as a change in isoform ratio $\geq$ 10%. For the LSK, MP, and K562 data, we used two-sample statistical comparisons (Wagenmakers' framework; Bayes factor $\geq$ 5);]. For the AML and CMML data, we used group statistical comparisons (Mann-Whitney *U* test, $p \leq 0.05$). Real-time PCR was used to measure *EZH2* cassette exon inclusion as described in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the RNA sequencing data reported in this paper is Gene Expression Omnibus (GEO): GSE65349.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.ccell.2015.04.006.

AUTHOR CONTRIBUTIONS

E.K., J.I., Y.L., G.M.D., F.H.-T.A., S.H., R.K.B, and O.A.-W. designed the study. E.K., S.L., Y.R.C., J.B.M., H.C., M.-K.K., and O.A.-W. performed animal experiments and generated mice. J.I., A.R., M.M., and A.S.Z. generated SRSF2 constructs, K562 cell lines, and CMML RNA-seq data. S.B. and P.S. provided additional SRSF2 constructs. R.K.B. performed the RNA-seq analysis. A.R., J.D., and O.A.-W. provided primary patient leukemia samples. C.L. and I.A. provided advice on animal experiments and helped generate RNA-seq data. S.A. and C.Y.P. performed cytopathologic and histopathologic analyses. Y.L., G.M.D., Y.L., Y.M., F.H.-T.A., and S.H. prepared protein and RNA samples for NMR and ITC studies and performed analyses. J.I. created minigenes and conducted splicing assays. E.K., G.M.D., F.H.-T.A., S.H., R.K.B., and O.A.-W. prepared the manuscript with help from all co-authors.

ACKNOWLEDGMENTS

E.K. is supported by the Worldwide Cancer Research Fund. A.R. was supported by the NIH/NHLBI (U01 HL099993), NIH/NIDDK (K08 DK082783), the J.P. McCarthy Foundation, and the Storb Foundation. S.H. and O.A.-W. are supported by grants from the Edward P. Evans Foundation. S.H. was supported by Yale Comprehensive Cancer Center institutional funds. R.K.B. was supported by the Hartwell Innovation Fund, Damon Runyon Cancer Research Foundation (DFS 04-12), Ellison Medical Foundation (AG-NS-1030-13), NIH/NIDDK (R56 DK103854), NIH/NCI recruitment support (P30 CA015704), and Fred Hutchinson Cancer Research Center institutional funds. J.O.I. was supported by an NIH/NCI training grant (T32 CA009657) and NIH/NIDDK pilot study (P30 DK056465). C.L. is supported by a career development award grant from the Leukemia and Lymphoma Society and an ATIP-Avenir grant from the French government. O.A.-W. is supported by an NIH K08 clinical investigator award (1K08CA160647-01), a Department of Defense Postdoctoral Fellow Award in Bone Marrow Failure Research (W81XWH-12-1-0041), the Josie Robertson Investigator Program, and a Damon Runyon Clinical Investigator Award with support from the Evans Foundation. F.H.-T.A. acknowledges support from the NCCR RNA and Disease funded by the Swiss National Science Foundation and the SNF Sinergia CRSII3_127454. Y.L. and Y.M. were supported by NIH/NIGMS grant R01 GM102869 and Senior Research Fellowship Grant 101908/Z/13/Z (to Y.M.) from the Wellcome Trust. J.D. acknowledges assistance from Dr. Nezih Cereb, HistoGenetics (Ossining, NY).

Received: December 2, 2014
Revised: February 19, 2015
Accepted: April 10, 2015
Published: May 11, 2015

REFERENCES

Bejar, R., Stevenson, K.E., Caughey, B.A., Abdel-Wahab, O., Steensma, D.P., Galili, N., Raza, A., Kantarjian, H., Levine, R.L., Neuberg, D., et al. (2012). Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. J. Clin. Oncol. *30*, 3376–3382.

Brooks, A.N., Choi, P.S., de Waal, L., Sharifnia, T., Imielinski, M., Saksena, G., Pedamallu, C.S., Sivachenko, A., Rosenberg, M., Chmielecki, J., et al. (2014). A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events. PLoS ONE *9*, e87361.

Daubner, G.M., Cléry, A., Jayne, S., Stevenin, J., and Allain, F.H. (2012). A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. EMBO J. *31*, 162–174.

Ernst, T., Chase, A.J., Score, J., Hidalgo-Curtis, C.E., Bryant, C., Jones, A.V., Waghorn, K., Zoi, K., Ross, F.M., Reiter, A., et al. (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat. Genet. *42*, 722–726.

Graubert, T.A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C.L., Shao, J., Krysiak, K., Harris, C.C., Koboldt, D.C., Larson, D.E., et al. (2012). Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat. Genet. *44*, 53–57.

Graveley, B.R., and Maniatis, T. (1998). Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. Mol. Cell *1*, 765–771.

Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., Schnittger, S., Sanada, M., Kon, A., Alpermann, T., et al. (2014). Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia *28*, 241–247.

Hubert, C.G., Bradley, R.K., Ding, Y., Toledo, C.M., Herman, J., Skutt-Kakaria, K., Girard, E.J., Davison, J., Berndt, J., Corrin, P., et al. (2013). Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. Genes Dev. *27*, 1032–1045.

Ilagan, J.O., Ramakrishnan, A., Hayes, B., Murphy, M.E., Zebari, A.S., Bradley, P., and Bradley, R.K. (2015). U2AF1 mutations alter splice site recognition in hematological malignancies. Genome Res. *25*, 14–26.

Kühn, R., Schwenk, F., Aguet, M., and Rajewsky, K. (1995). Inducible gene targeting in mice. Science *269*, 1427–1429.

Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., and Brenner, S.E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature *446*, 926–929.

Liu, H.X., Chew, S.L., Cartegni, L., Zhang, M.Q., and Krainer, A.R. (2000). Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. Mol. Cell. Biol. *20*, 1063–1071.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X., et al. (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell *20*, 11–24.

Muto, T., Sashida, G., Oshima, M., Wendt, G.R., Mochizuki-Kashio, M., Nagata, Y., Sanada, M., Miyagi, S., Saraya, A., Kamio, A., et al. (2013). Concurrent loss of Ezh2 and Tet2 cooperates in the pathogenesis of myelodysplastic disorders. J. Exp. Med. *210*, 2627–2639.

Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A., Blume, J.E., and Ares, M., Jr. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev. *21*, 708–718.

Nikoloski, G., Langemeijer, S.M., Kuiper, R.P., Knops, R., Massop, M., Tönnissen, E.R., van der Heijden, A., Scheele, T.N., Vandenberghe, P., de Witte, T., et al. (2010). Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. Nat. Genet. *42*, 665–667.

Oguro, H., Ding, L., and Morrison, S.J. (2013). SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. Cell Stem Cell *13*, 102–116.

Papaemmanuil, E., Cazzola, M., Boultwood, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J.S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C., et al.; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium (2011). Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N. Engl. J. Med. *365*, 1384–1395.

Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C.J., Ellis, P., Wedge, D.C., Pellagatti, A., et al.; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood *122*, 3616–3627, quiz 3699.

Pronk, C.J., Rossi, D.J., Månsson, R., Attema, J.L., Norddahl, G.L., Chan, C.K., Sigvardsson, M., Weissman, I.L., and Bryder, D. (2007). Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. Cell Stem Cell *1*, 428–442.

Przychodzen, B., Jerez, A., Guinta, K., Sekeres, M.A., Padgett, R., Maciejewski, J.P., and Makishima, H. (2013). Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. Blood *122*, 999–1006.

Quesada, V., Conde, L., Villamor, N., Ordóñez, G.R., Jares, P., Bassaganyas, L., Ramsay, A.J., Beà, S., Pinyol, M., Martínez-Trillos, A., et al. (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat. Genet. *44*, 47–52.

Schaal, T.D., and Maniatis, T. (1999). Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Mol. Cell. Biol. *19*, 261–273.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. *15*, 1034–1050.

’t Hoen, P.A., Hirsch, M., de Meijer, E.J., de Menezes, R.X., van Ommen, G.J., and den Dunnen, J.T. (2011). mRNA degradation controls differentiation state-dependent differences in transcript and splice variant abundance. Nucleic Acids Res. *39*, 556–566.

Vannucchi, A.M., Lasho, T.L., Guglielmelli, P., Biamonte, F., Pardanani, A., Pereira, A., Finke, C., Score, J., Gangat, N., Mannarelli, C., et al. (2013). Mutations and prognosis in primary myelofibrosis. Leukemia *27*, 1861–1869.

Wagenmakers, E.J., Lodewyckx, T., Kuriyal, H., and Grasman, R. (2010). Bayesian hypothesis testing for psychologists: a tutorial on the Savage-Dickey method. Cognit. Psychol. *60*, 158–189.

Wang, H.Y., Xu, X., Ding, J.H., Bermingham, J.R., Jr., and Fu, X.D. (2001). SC35 plays a role in T cell development and alternative splicing of CD45. Mol. Cell *7*, 331–342.

Wang, L., Lawrence, M.S., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K., Werner, L., Sivachenko, A., DeLuca, D.S., Zhang, L., et al. (2011). SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N. Engl. J. Med. *365*, 2497–2506.

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., et al. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64–69.

Zahler, A.M., Damgaard, C.K., Kjems, J., and Caputi, M. (2004). SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. J. Biol. Chem. *279*, 10077–10084.

Zhang, S.J., Rampal, R., Manshouri, T., Patel, J., Mensah, N., Kayserian, A., Hricik, T., Heguy, A., Hedvat, C., Gönen, M., et al. (2012). Genetic analysis of patients with leukemic transformation of myeloproliferative neoplasms shows recurrent SRSF2 mutations that are associated with adverse outcome. Blood *119*, 4480–4485.

Specific molecular signatures predict decitabine response in chronic myelomonocytic leukemia

Kristen Meldi,[1] Tingting Qin,[1] Francesca Buchi,[2] Nathalie Droin,[3] Jason Sotzen,[1] Jean-Baptiste Micol,[3,4] Dorothée Selimoglu-Buet,[3] Erico Masala,[2] Bernardino Allione,[5,6] Daniela Gioia,[6,7] Antonella Poloni,[6,8] Monia Lunghi,[6,9] Eric Solary,[3] Omar Abdel-Wahab,[4] Valeria Santini,[2,6] and Maria E. Figueroa[1]

[1]University of Michigan Medical School, Department of Pathology, Ann Arbor, Michigan, USA. [2]University of Florence, AOU Careggi, Hematology, Florence, Italy. [3]INSERM U1009, Gustave Roussy Cancer Center, Villejuif, France. [4]Memorial Sloan Kettering Cancer Center, Human Oncology and Pathogenesis Program and Leukemia Service, New York, New York, USA. [5]Hematology, Ospedale S. Giovanni Battista Molinette, Torino, Italy. [6]Fondazione Italiana per le Sindromi Mielodisplastiche (FISM), Alessandria, Italy. [7]Azienda Ospedaliera SS. Antonio e Biagio e C. Arrigo, Alessandria, Italy. [8]Hematology, Università Politecnica delle Marche, Ancona, Italy. [9]Department of Hematology, Amedeo Avogadro University, Novara, Italy.

Myelodysplastic syndromes and chronic myelomonocytic leukemia (CMML) are characterized by mutations in genes encoding epigenetic modifiers and aberrant DNA methylation. DNA methyltransferase inhibitors (DMTis) are used to treat these disorders, but response is highly variable, with few means to predict which patients will benefit. Here, we examined baseline differences in mutations, DNA methylation, and gene expression in 40 CMML patients who were responsive or resistant to decitabine (DAC) in order to develop a molecular means of predicting response at diagnosis. While somatic mutations did not differentiate responders from nonresponders, we identified 167 differentially methylated regions (DMRs) of DNA at baseline that distinguished responders from nonresponders using next-generation sequencing. These DMRs were primarily localized to nonpromoter regions and overlapped with distal regulatory enhancers. Using the methylation profiles, we developed an epigenetic classifier that accurately predicted DAC response at the time of diagnosis. Transcriptional analysis revealed differences in gene expression at diagnosis between responders and nonresponders. In responders, the upregulated genes included those that are associated with the cell cycle, potentially contributing to effective DAC incorporation. Treatment with CXCL4 and CXCL7, which were overexpressed in nonresponders, blocked DAC effects in isolated normal $CD34^+$ and primary CMML cells, suggesting that their upregulation contributes to primary DAC resistance.

Introduction

Chronic myelomonocytic leukemia (CMML) is a myelodysplastic syndrome/myeloproliferative neoplasm (MDS/MPN) overlap syndrome (1) that was historically classified within MDS (2) until 2001 (3). CMML shares many characteristics with MDS, including dysplasia in one or more myeloid cell lineages and increased risk of transformation into acute myeloid leukemia (AML). However, a distinguishing feature of CMML is the presence of persistent peripheral monocytosis ($>1 \times 10^9$/l). CMML can be subdivided into 2 subtypes on the basis of blast count: CMML1, with less than 10% bone marrow (BM) blasts, and CMML2, which has between 10% and 19% blasts.

Substantial epigenetic abnormalities have been described in both MDS and MDS/MPN. Mutations in epigenome-modifying enzymes are highly prevalent in these disorders, including those responsible for DNA methylation and demethylation — DNA methyltransferase 3A (DNMT3A) (4) and ten-eleven translocation 2 (TET2) (5, 6), respectively — as well as those involved in histone-modifying complexes — additional sex combs-like 1 (ASXL1) (7) and enhancer of zeste homolog 2 (EZH2) (8–11). Although the precise mechanisms through which these mutations drive the aberrant epigenetic changes observed in MDS are still not completely understood, it has been shown that MDS and MDS/MPN are characterized by a DNA hypermethylation that increases with disease severity (12, 13).

MDS and MDS/MPN are resistant to conventional chemotherapies; however, epigenome-modifying drugs can be used successfully as therapeutics to treat these disorders. In particular, the nucleoside analogs azacytidine (AZA) and decitabine (DAC) are commonly used to treat MDS and CMML (14, 15). Both AZA and DAC are DNA methyltransferase inhibitors (DMTis), and while their precise mechanism of action in treating MDS and MDS/MPN remains a point of controversy, they are known to be incorporated into DNA during the S phase, where they covalently trap DNA methyltransferases and target them for proteasomal degradation (16, 17). DMTis can also cause DNA damage (18), and because AZA is mostly incorporated into RNA, it may have additional effects on RNA processing and translation (19). Despite the utility of DAC and AZA, only a subset of MDS and CMML patients respond to them. Only approximately 50% of patients treated with DMTis show a hematological improvement (HI) or better that is associated with a survival benefit (20). Furthermore, as many as 6 months of treatment may be required for the therapeutic benefit of DMTis to become apparent, thus forcing half of the patients

Authorship note: Kristen Meldi, Tingting Qin, Valeria Santini, and Maria E. Figueroa contributed equally to this work.
Conflict of interest: The authors have declared that no conflict of interest exists.
Submitted: August 27, 2014; **Accepted:** February 9, 2015.
Reference information: *J Clin Invest.* 2015;125(5):1857–1872. doi:10.1172/JCI78752.

Table 1. Clinical characteristics of the FISM CMML patient cohort treated with DAC

Clinical characteristics	Responders	Nonresponders	*P* value
Total no. of patients	20	20	
CMML1, no. (%)	15 (75%)	10 (50%)	NS[A]
CMML2, no. (%)	5 (25%)	10 (50%)	
Male, no. (%)	14 (70%)	14 (70%)	NS[A]
Female, no. (%)	6 (30%)	6 (30%)	
Median age, yr (range)	73.5 (45–84)	70.5 (41–82)	NS[B]
Median survival, mo (range)	26.5 (6–39)	13.5 (2–25)	$P = 0.0004$[C]
Median hemoglobin, no. (range)	10 (7.2–14.9)	9.7 (6.6–13.8)	NS[A]
Median marrow blasts, % (range)	5 (0–18)	7 (0–19)	NS[D]
Median monocytes, % (range)	24 (2–67)	22 (5–45)	NS[D]
Median wbc, % (range)	17.8 (3.7–75.2)	18.9 (2.8–52.5)	NS[A]
Cytogenetics			
Normal	14	14	NS[A]
Abnormal	6	6	
Splenomegaly	9	7	NS[A]
Hepatomegaly	8	5	NS[A]
Lymphadenomegaly	2	3	NS[A]

[A]Fisher's exact test; [B]Student's *t* test; [C]log-rank test; [D]Wilcoxon rank-sum test.

to undergo long periods of treatment before they can be deemed resistant to this therapy. Currently, there are very few means of predicting response versus resistance, and even this is exclusive to AZA (21). Additionally, few alternative treatments exist for patients who fail to respond to DMTis, and their prognosis is extremely poor. Therefore, it is critical that we better understand the molecular profiles associated with sensitivity and resistance to DMTis in order to improve risk stratification strategies as well as shed light on the mechanisms of resistance.

While some studies have suggested that reversal of methylation and/or transcript reexpression of certain loci was associated with clinical response to DMTis (22–28), epigenetic studies to date have failed to identify any strong correlation between response to these agents and the presence of specific baseline DNA methylation profiles (23, 26, 27, 29, 30). We hypothesized that this lack of correlation was due to the promoter-centric nature of assays used over the past decade and that methylation differences associated with potential for therapeutic response were likely present in these patients upon diagnosis at promoter-distal and intergenic regulatory regions. In this study, we report, for the first time to our knowledge, the identification of DNA methylation and expression differences in diagnostic BM specimens from a cohort of CMML patients treated with DAC. These differences, detected through the use of genome-wide next-generation sequencing assays, reveal underlying biological differences between these 2 groups of patients and point to a novel mechanism of resistance to DMTis.

Results

Somatic mutations do not correlate with response to DAC in CMML. Somatic mutations in epigenome-modifying enzymes and other genes are prevalent in MDS and CMML (4–6, 31–35). Recently, it has been reported that mutations in *TET2* and *DNMT3A* are associated with improved response to DMTi therapy in MDS and related disorders (36–38). Despite this, the presence of these mutations did not translate to an improved overall survival rate in any of these studies, indicating that therapeutic response and survival benefit are likely influenced by multiple different factors. Moreover, these findings have not been recapitulated in CMML exclusively (39). To determine whether particular genetic or epigenetic abnormalities are associated with DMTi sensitivity or resistance in this disease, we studied a cohort of primary CMML cases. BM mononuclear cells (BM MNCs) were collected from 40 patients with de novo CMML at the time of their diagnosis. All patients included in this study were enrolled in a clinical trial conducted by the FISM and received single-agent treatment with DAC as frontline therapy (20 mg/m^2/day for 5 days), and response was evaluated after 6 cycles of treatment. Responsive patients ($n = 20$) were defined as those who achieved either complete remission, marrow complete remission, partial remission, or HI, as defined by the 2006 International Working Group (IWG) response criteria for myelodysplasia (40). Patients with either stable disease or progressive disease were considered to have primary resistance to DAC ($n = 20$). As shown in Table 1, there were no significant differences in terms of age, gender, BM monocytosis, blast percentage, cytogenetics, or presence of either splenomegaly or extramedullary lesions between responder and nonresponder patients. Using MiSeq to sequence DNA isolated from the diagnostic BM MNCs, we performed targeted resequencing of the following panel of genes mutated at frequencies greater than 5% in CMML: *SRSF2*, *TET2*, *ASXL1*, *NRAS*, *DNMT3A*, *RUNX1*, *U2AF1*, *TP53*, *JAK2*, *KIT*, *KRAS*, *SF3B1*, *EZH2*, *IDH1*, and *IDH2*. As with previous reports, *SRSF2*, *TET2*, and *ASXL1* were the most frequently mutated genes in this cohort of patients (6, 32, 34, 35, 41–44). However, no somatic mutation was significantly correlated with response to DAC in our cohort (Fisher's exact test, P = NS for all mutations) (Figure 1A and Table 2).

We have previously shown, as have others, that distinct DNA methylation profiles in AML and acute lymphoid leukemia (ALL) are strongly correlated with the presence of specific molecular and cytogenetic subtypes (12, 45–48). To determine whether similarly distinct methylation patterns in CMML can be linked to the presence of specific somatic mutations, we examined DNA methylation patterns in the same specimens through enhanced reduced representation bisulfite sequencing (ERRBS) (45), a deep-sequencing method that captures and accurately quantifies DNA methylation at approximately 3 million CpG sites. ERRBS data were available for 39 of the 40 patients (19 nonresponders and 20 responders). The percentage of methylation measured by ERRBS was highly concordant with the findings of the quantitative single-locus DNA methylation validation assay MassARRAY EpiTYPER (ref. 49 and Supplemental Figure 1; supplemental material available online with this article; doi:10.1172/JCI78752DS1). Unsupervised clustering analysis of the patients based on their DNA methylation patterns did not reveal a correlation between gene mutations and particular methylation clusters (Figure 1B). In addition, there was no significant difference in the observed

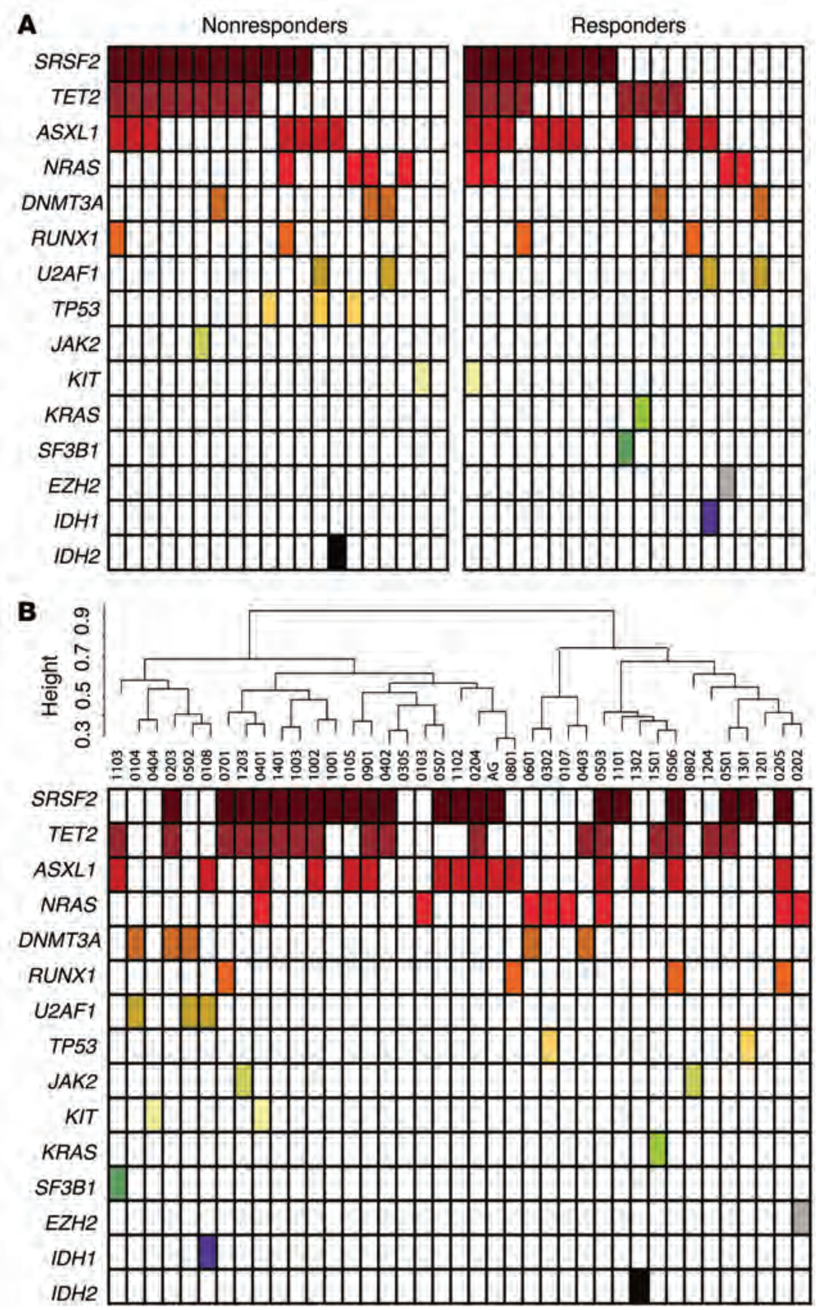


Figure 1. Somatic mutations in CMML do not correlate with DAC response or specific epigenetic clusters. Mutational status of a panel of 15 genes frequently mutated in CMML according to (**A**) therapeutic response to DAC or (**B**) DNA methylation hierarchical clustering.

patient survival time between the 2 top-level methylation clusters (log-rank test, $P = 0.33$).

Next, we performed supervised analyses comparing *TET2*, *ASXL1*, *DNMT3A*, and *SRSF2* WT and mutant cases to identify the differentially methylated regions (DMRs) associated with each of these mutations. As expected, given its role in de novo DNA methylation, we identified a predominantly hypomethylated profile associated with *DNMT3A* mutations (total DMRs: 243; hypomethylated DMRs [hypo-DMRs]: 197; hypermethylated DMRs [hyper-DMRs]: 46) that was targeted mainly at intergenic and intronic regions (Figure 2A). By contrast, *TET2* loss-of-function mutations were associated with the presence of hypermethylation compared with that seen in *TET2* WT cases (total DMRs: 188; hypo-DMRs: 48; hyper-DMRs: 140) (Figure 2B). Mutations in *ASXL1,* another epigenetic modifier, were associated with a specific signature consisting of equal proportions of hyper- and hypo-DMRs (total DMRs: 144, hypo-DMRs: 82, hyper-DMRs: 62). Both hyper- and hypo-DMRs in *ASXL1*-mutant CMML cases were strongly depleted at promoter regions (hyper-DMRs 3% vs. background 21%, $P = 6.79 \times 10^{-5}$; hypo-DMRs 5% vs. background 21%, $P = 4.30 \times 10^{-5}$) and significantly enriched at intergenic regions (hypo-DMRs 54% vs. background 38%, $P = 2.84 \times 10^{-3}$) (Figure 2C). Notably, mutations in the splicing factor *SRSF2* were linked to the strongest DNA methylation differences, with a total of 724 DMRs (hypo-DMRs: 383; hyper-DMRs: 341). In this case, hypermethylated DMRs were strongly enriched at promoter regions (hyper-DMRs 31% vs. background 21%, $P = 1.44 \times 10^{-5}$) and depleted at introns (hyper-DMRs 19% vs. background 33%, $P = 1.50 \times 10^{-8}$) (Figure 2D). While *SRSF2* itself does not have any direct epigenetic function, it is likely that mutations in this gene lead to mis-splicing and the consequent deregulation of other epigenome-modifying genes, resulting in this strong epigenetic signature. Additionally, the observed survival time was not significantly different between the patients with or without individual *DNMT3A*, *TET2*, *ASXL1*, and *SRSF2* mutations (log-rank test, $P = 0.61$, 0.067, 0.93, and 0.58, respectively).

A specific epigenetic profile distinguishes DAC-resistant CMML patients at diagnosis. Previous efforts by many groups, including ours, have failed to identify baseline epigenetic differences between DMTi-sensitive and -resistant patients (12, 27, 30). However, all of these studies were performed using platforms that examined DNA methylation within CpG islands and gene promoters. A growing body of recent evidence suggests that DNA methylation and other epigenetic modifications at enhancers and other distal regulatory regions play a key role in transcriptional regulation and that these regions are often located at a significant distance from the transcription start site of the target gene (50). Therefore, we hypothesized that key epigenetic differences may exist between DAC-sensitive and -resistant patients at diagnosis that are located distally from promoters, targeting enhancers and other distal regulatory regions.

For this purpose, we used the ERRBS assay, a deep-sequencing–based method that targets not only promoter regions but also intronic, exonic, and distal intergenic regions (45). Using the MethylSig package, we performed a direct comparison between the diagnostic DNA methylation profiles of DAC-sensitive and DAC-resistant patients (51). We identified 167 DMRs that displayed a methylation difference of 25% or more between respond-

Table 2. Somatic mutations of the FISM cohort did not correlate with response

Mutation	Nonresponders (n = 20)	Responders (n = 20)	Total (n = 40)	P value[A]
SRSF2	60.0% n = 12	45.0% n = 9	52.5% n = 21	0.53
TET2	45.0% n = 9	40.0% n = 8	42.5% n = 17	1.0
ASXL1	35.0% n = 7	45.0% n = 9	40.0% n = 16	0.75
NRAS	20.0% n = 4	20.0% n = 4	20.0% n = 8	1.0
DNMT3A	15.0% n = 3	10.0% n = 2	12.5% n = 5	1.0
RUNX1	10.0% n = 2	10.0% n = 2	10.0% n = 4	1.0
U2AF1	10.0% n = 2	10.0% n = 2	10.0% n = 4	1.0
TP53	15.0% n = 3	0.0% n = 0	7.5% n = 3	0.23
JAK2	5.0% n = 1	5.0% n = 1	5.0% n = 2	1.0
KIT	5.0% n = 1	5.0% n = 1	5.0% n = 2	1.0
KRAS	0.0% n = 0	5.0% n = 1	2.5% n = 1	1.0
SF3B1	0.0% n = 0	5.0% n = 1	2.5% n = 1	1.0
EZH2	0.0% n = 0	5.0% n = 1	2.5% n = 1	1.0
IDH1	0.0% n = 0	5.0% n = 1	2.5% n = 1	1.0
IDH2	5.0% n = 1	0.0% n = 0	2.5% n = 1	1.0

[A]Fisher's exact test.

ers and nonresponders and that were statistically significant at an FDR of less than 0.1. Among these DMRs were regions displaying higher methylation in responders, as well as regions of lower methylation as compared with those in nonresponders (Figure 3A and Supplemental Table 1). Hierarchical clustering of our cohort using these DMRs was sufficient to achieve a perfect segregation of DAC-sensitive and -resistant patients (Figure 3B). These findings indicate that numerous epigenetic differences exist at the time of diagnosis that correlate with a patient's likelihood of responding to DAC treatment.

Response-associated DMRs localize preferentially to distal regulatory regions. Next, we sought to determine whether DMRs were distributed evenly across the genome or whether they were enriched at specific genomic regions. For this, we analyzed both the genomic distribution of DMRs as well as their association with known regulatory regions. Notably, our analysis of the distribution of DMRs relative to coding regions revealed that DMRs were significantly depleted at promoter regions (DMRs 10% vs. background 21%, binomial test $P = 6.70 \times 10^{-5}$), with a concurrent enrichment at intronic regions, thus confirming our initial hypothesis. This distribution was not the same across hyper- and hypo-DMRs. While all DMRs were depleted at promoter regions, hyper-DMRs were significantly enriched at introns (hyper-DMRs 49% vs. background 33%, binomial test $P = 1.29 \times 10^{-3}$), while hypo-DMRs were enriched at intergenic regions (hypo-DMRs 49% vs. 38% background, binomial test $P = 0.03$) (Figure 4A).

Next, we sought to determine the association of DMRs with regulatory regions. For this purpose, we analyzed their relative enrichment at CpG island and enhancer regions. Analysis of CpG islands and CpG shores demonstrated that DMRs were also significantly depleted at CpG islands (DMRs 14% vs. background 25%, binomial test $P = 2.8 \times 10^{-4}$), with enrichment at CpG shores (DMRs 22% vs. background 15%, binomial test $P = 8.79 \times 10^{-3}$). This pattern was conserved across both hyper- and hypo-DMRs (Figure 4B).

Recently, DNA methylation at enhancers was reported to strongly correlate with aberrant gene expression observed in cancer cells (52). We hypothesized that differential DNA methylation at enhancers, rather than at promoters, may be better correlated with differential response to DAC in CMML. Enrichment analysis of all DMRs relative to intragenic and intergenic enhancers revealed that DMRs were enriched for intragenic enhancers (DMRs 25% vs. background 18%, binomial test $P = 0.01$). When this analysis was stratified into hyper- and hypo-DMRs, it became apparent that hyper-DMRs showed the strongest enrichment at enhancer regions and, in particular, at enhancers located within gene bodies (hyper-DMRs 32% vs. background 18%, binomial test $P = 8.14 \times 10^{-4}$). Conversely, hypo-DMRs were not significantly enriched at enhancer regions and were similarly distributed within gene body and intergenic enhancers (Figure 4C).

Finally, we asked whether the DMRs associated with DAC response were specifically enriched within relevant biological pathways. The 167 DMRs were annotated to known genes, and pathway enrichment analysis was performed against the KEGG pathway database. The MAPK signaling pathway, which plays a key role in the cell cycle, apoptosis, cell proliferation, and differentiation, was significantly enriched in DMR-associated genes (hypergeometric test $P = 7.68 \times 10^{-3}$, FDR = 0.084) (Supplemental Figure 2A). There were 7 DMRs that were annotated to MAPK pathway genes, including *STMN1*, *CACNAE1*, *PRKCB*, *MAPT*, *NFATC1*, *CRKL*, and *MKNK2* (Supplemental Table 2). Three of these DMRS — those annotated to *STMN1*, *CACNAE1*, and *MAPT* — were hypermethylated in DAC nonresponders, while *MKNK2*-, *NFATC1*-, *CRKL*-, and *PRKCB*-associated DMRs were hypermethylated in DAC responders (summarized in Supplemental Table 2). To further validate epigenetic deregulation of the MAPK signaling pathway in these patients, we performed MassARRAY EpiTYPER analysis of 3 of the affected MAPK genes in the pathway in a subset of samples (Supplemental Figure 2B). This analysis confirmed the increased methylation in the *STMN1* and *CACNAE1* DMRs in nonresponder patients, as well as validated the increased methylation of the *NFATC1* DMR in responders.

DNA methylation differences can be harnessed for therapeutic response prediction. Given that our data identified, for the first time to our knowledge, the existence of baseline DNA methylation differences between DAC responders and nonresponders prior to DAC treatment, we hypothesized that these unique methylation profiles could be harnessed to predict at the time of diagnosis which patients would be sensitive or resistant to treatment. To test this, we used the percentage of cytosine methylation at each genomic location among patients in the FISM cohort (cohort 1) as potential predictors and applied a machine-learning approach, support vector machine (SVM) (53), to build a classifier (see details in Methods). Twenty-one 25-bp tile regions were identified by feature selection as the predictors with the highest predictability in the SVM classifier (Figure 5A, Supplemental Figure 3A, and Supplemental Table 3). Unsupervised analysis using only the methylation levels at the 21 selected tile regions revealed that they were sufficient to almost

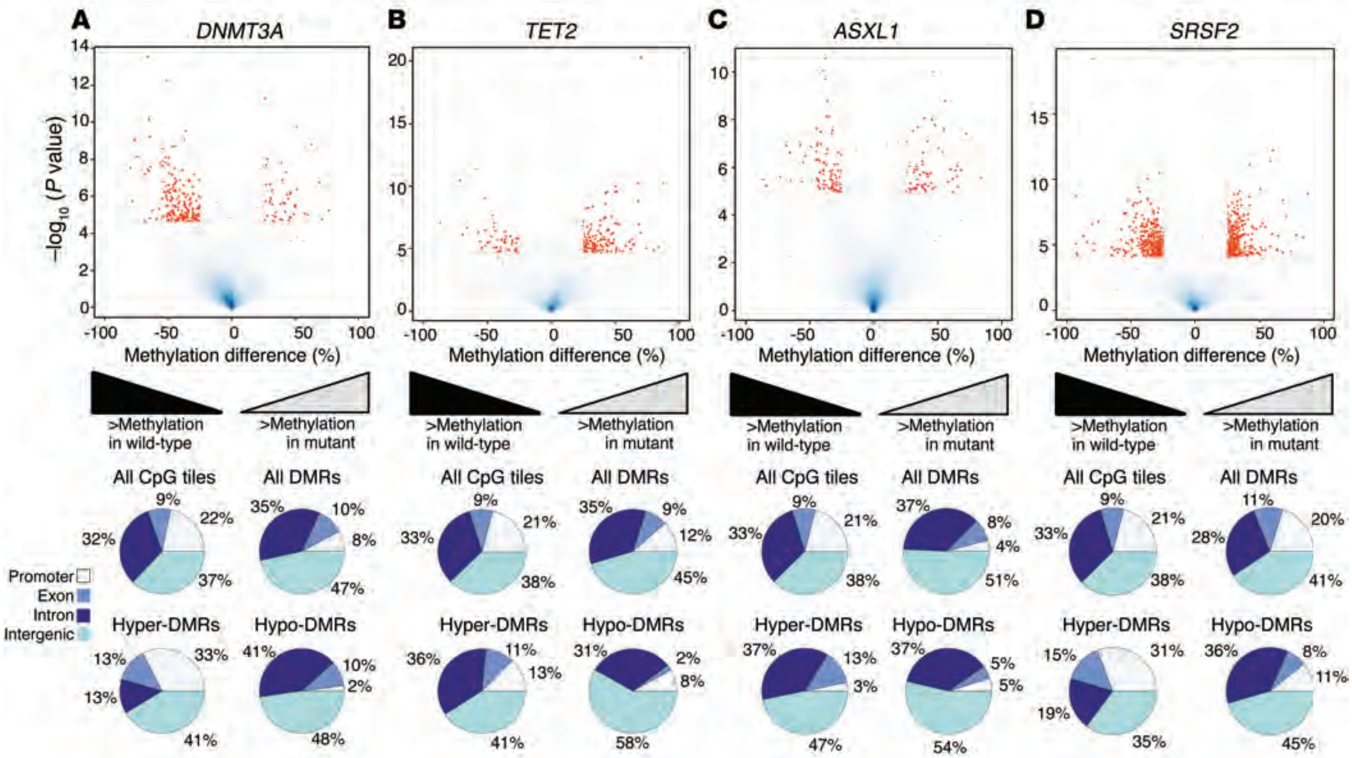


Figure 2. Distinct DNA methylation profiles are associated with recurrent somatic mutations in *DNMT3A*, *TET2*, *ASXL1*, and *SRSF2*. Volcano plots illustrating the methylation differences between *DNMT3A*-mutant (n = 5) (**A**), *TET2*-mutant (n = 17) (**B**), *ASXL1*-mutant (n = 15) (**C**), or *SRSF2*-mutant (n = 21) (**D**) samples versus WT patients (n = 39 for the number of mutated samples). DMRs are indicated by red dots (beta-binomial test, FDR <0.1 and absolute methylation different ≥25%). Pie charts illustrate the relative proportion of CpG tiles and DMRs annotated to the RefSeq promoter, exonic, intronic, and intergenic regions.

separate the 39 samples by response (Figure 5B and Supplemental Figure 3, B and C). There was no defined clustering of the patients according to their specific degree of response as shown by multidimensional scaling (MDS) analysis (Supplemental Figure 3C), which is concordant with the fact that the classifier was built to identify an all-or-nothing response versus no response and not to distinguish between types of responses. Ten-fold cross-validation was performed using the cases from cohort 1 to evaluate the predictive performance of the classifier, and the reported area under the receiver operating characteristic curve (ROC-AUC) was 0.99, indicating a strong predictive accuracy for the classifier model (Supplemental Figure 3D). In order to further assess the robustness of the SVM classifier built with the 21 selected features, we performed 3 different random splits of the same cohort 1 into training and test sets. We trained the classifier on each of the 3 sets of randomly selected samples and predicted the responses for the remaining samples in the cohort. The classifier was able to accurately predict response to DAC in 18 of 19 (accuracy = 94.74%) (Table 3), 13 of 14 (accuracy = 92.86%), and 9 of 9 (accuracy = 100%) patients, respectively (Supplemental Figure 4A).

Since validation in an independent cohort of patients is the gold standard for biomarker development, we identified a second cohort of patients in which to test the performance of our SVM classifier. Twenty-eight additional diagnostic CMML specimens from patients enrolled in a clinical trial from the Groupe Francophone des Myelodysplasies (GFM), all of whom had been treated with the same DAC regimen of 20 mg/m^2/day for 5 days, were collected and subjected to ERRBS (Table 4 and Supplemental Table 4). Specimens from this second cohort (cohort 2) of 12 responder and 16 nonresponder patients consisted of sorted monocytes from peripheral blood (PB). The SVM classifier that had been developed using cohort 1 was applied blindly to these samples, without any prior knowledge of the therapeutic response labels for this second cohort. Due to the stochastic nature of ERRBS, the CpG coverage is never identical across all samples, thus leading to missing values for some regions of interest. In effect, only 6 of the 21 features were present in all 28 samples in cohort 2. Therefore, using only these 6 features, we first trained our SVM classifier on the 39 samples of the FISM cohort (cohort 1) and then applied the trained classifier on the GFM cohort (cohort 2). As shown in Table 5 and Supplemental Figure 4B, despite this limitation, the 6-feature classifier was still capable of correctly predicting response in 20 of 28 patients in the GFM cohort (accuracy = 71% and AUC = 0.82). Next, in order to increase the number of features being tested while still retaining a large enough cohort in which to test the predictive accuracy, we used 14 of the 21 features of the SVM classifier to predict response for 19 patients in the GFM cohort. Once again, we used only these 14 features to train the model on cohort 1, which consisted of the initial 39 patients, and then blindly applied the model to the 19 test samples from the GFM cohort (cohort 2). This modified classifier with 14 features was capable of accurately predicting therapeutic outcome for 15 of the 19 patients, which represents an accuracy of 79% and an AUC of 0.83. (Table 5 and Supplemental Figure 4B). Finally, we determined that of the original 21 features, 16 was the

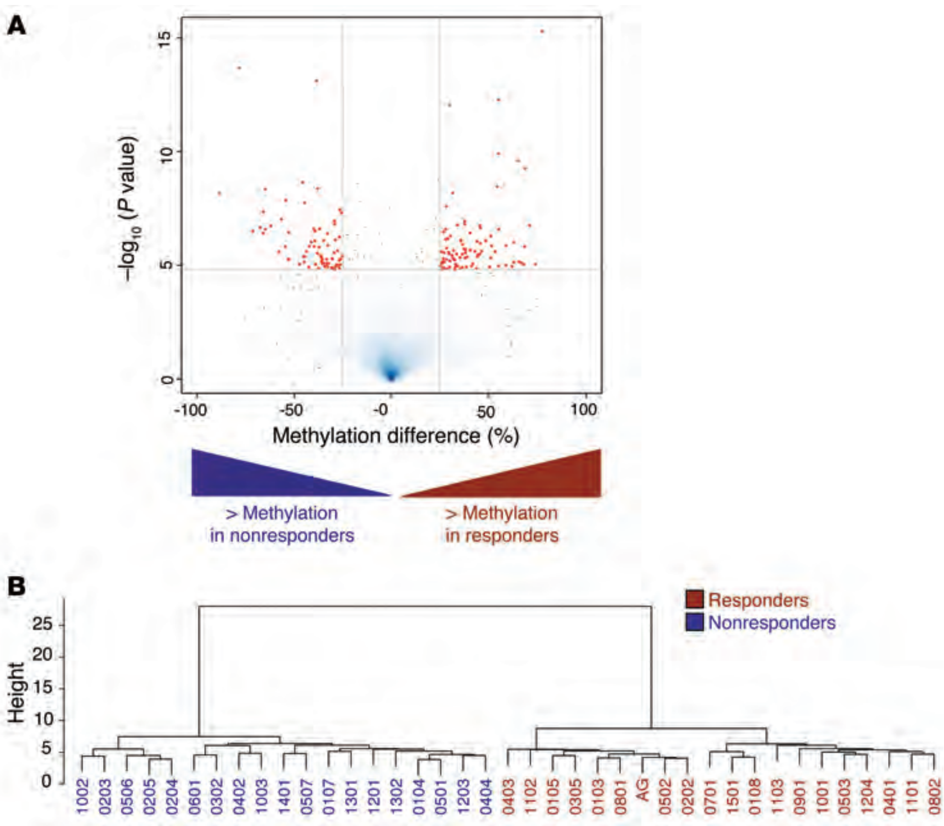


Figure 3. Baseline DNA methylation differences distinguish DAC responders and nonresponders at the time of diagnosis. (**A**) Volcano plot illustrating methylation differences between 20 DAC-sensitive and 19 DAC-resistant patients. Mean methylation difference between the 2 groups is represented on the x axis and statistical significance ($-\log_{10}$ P value) on the y axis. Beta-binomial test identified 167 DMRs, which are indicated by red dots (FDR <0.1 and absolute methylation difference ≥25%). (**B**) Hiearchical clustering of the patients using the 167 DMRs illustrates the power of these genomic regions in segregating the patients into nonresponder (blue) and responder (red) groups.

maximum number of features shared by at least 15 of the cohort-2 patients. We trained the model on cohort 1 using only these 16 shared features and then predicted response for the 15 patients in the independent cohort 2, achieving an accuracy of 87% with an AUC of 0.94 (Table 5 and Supplemental Figure 4B). These findings demonstrate that the SVM classifier developed using the original FISM cohort is general enough to be applied to and accurately predict the therapeutic outcome of fully independent samples (i.e., GFM cohort 2), which is a critical step in the development of a biomarker. Moreover, this robustness was maintained even across different cell types (BM MNCs in cohort 1 vs. PB monocytes in the validation cohort 2), further underscoring the power of the classifier to predict outcome in an independent cohort. While further validation in larger cohorts will be required to fully assess the accuracy of the features reported here, and additional studies of larger cohorts might help refine the selection of features to include those with the strongest accuracy over a large number of patients, our findings demonstrate that the epigenetic differences between responders and nonresponders at diagnosis have the potential to be harnessed as classifiers to predict clinical response to DAC.

DAC sensitivity can be linked to a specific transcriptional program at diagnosis. While it has been previously shown that reduced expression of uridine-cytidine kinase, an enzyme involved in nucleoside metabolism, is associated with response to AZA in MDS (54), we did not find that differential expression of this or other DMTi-metabolizing enzymes was associated with response to DAC in CMML (data not shown). Therefore, we sought to determine whether other transcriptional differences between DAC responders and nonresponders are indicative of response and can provide insight on functional pathways that contribute to DAC resistance. We performed RNA-sequencing (RNA-seq) on samples from 14 patients (8 responders and 6 nonresponders) in the cohort of CMML patients for whom we had high-quality RNA. Prior to performing differential analysis, we validated the ability of our RNA-seq approach to accurately detect quantitative variability by performing quantitative reverse transcriptase PCR (qRT-PCR) on RNAs from 13 of the 14 patients and determining the degree of agreement between the 2 methods (r = 0.85, R^2 value = 0.73, P < 0.0001) (Supplemental Figure 5A). As shown in Figure 6A, a direct comparison of the 2 groups of patients identified 601 genes with an absolute $\log_2$ fold change greater than 1 and a P value of less than 0.05. Notably, this gene signature consisted of a majority of genes overexpressed in DAC-sensitive patients (405 upregulated genes), with only a small proportion of genes downregulated in these patients (Supplemental Table 5).

In order to identify biological differences that might explain the difference between these patients in their therapeutic response to DAC, we performed gene set enrichment analysis (GSEA) (55). Gene sets enriched in DAC-sensitive patients at an FDR of less than 0.1 were involved in proliferation, cell cycle activity, and DNA replication (Figure 6B). Likewise, genes reported as being downregulated in quiescent versus dividing $CD34^+$ cells (56) were found to be upregulated in DAC responders. This enrichment of gene sets involved in the cell cycle and in DNA replication in DAC-sensitive patients is consistent with the requirement for DAC incorporation into the DNA during the S phase.

Primary resistance to DAC is associated with overexpression of ITGβ3 and the chemokines CXCL4 and CXCL7. As mentioned above, only a small fraction of genes were found to have at least

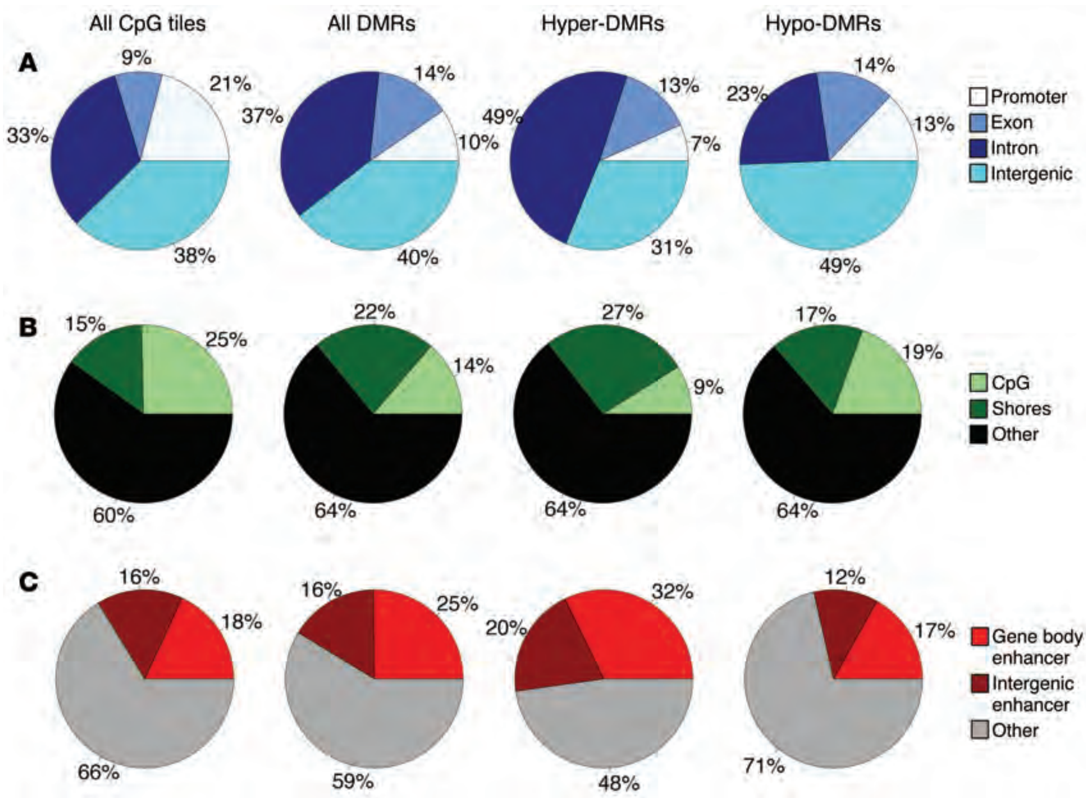


Figure 4. DMRs are enriched at distal intergenic regions and enhancers. (**A**) Pie charts illustrate the relative proportion of CpG tiles and DMRs annotated to RefSeq promoter, exonic, intronic, and intergenic regions. (**B**) Pie charts illustrate the relative proportion of CpG tiles and DMRs annotated to CpG islands, CpG shores, and regions beyond CpG shores. (**C**) Pie charts illustrate the relative proportion of CpG tiles and DMRs annotated to enhancers within gene bodies, enhancers within intergenic regions, and nonenhancer regions.

a 2-fold overexpression in DAC-resistant patients. Among these, 3 genes that have previously been implicated in chemoresistance and leukemogenesis were overexpressed in nonresponders: *CXCL4* (also known as *PF4*), *CXCL7* (also known as *PPBP*), and integrin β3 (*ITGB3*) (Figure 6C). Thus, we hypothesized that overexpression of these genes might be a potential mechanism through which CMML acquires resistance to DAC. First, as shown in Figure 7A, we validated the overexpression of these genes in DAC-resistant patients by qRT-PCR. Notably, there was a statistically significant linear correlation between the levels of CXCL4 and CXCL7 expression by both RNA-seq ($r = 0.9350$, $R^2 = 0.87$, $P < 0.0001$) and qRT-PCR ($r = 0.9865$, $R^2 = 0.9731$, $P < 0.0001$), suggesting that these factors act in concert in the BM microenvironment (Figure 7B). While both chemokines were originally thought to be produced exclusively by megakaryocytes, there is evidence that monocytes (57, 58) and other cells within the BM also produce CXCL4 and CXCL7 (refs. 59, 60, and Supplemental Figure 5, B and C). To further confirm the overexpression of these chemokines in nonresponder patients as well as to determine the cellular source and localization of the proteins in the BM, IHC was performed on a subset of paraffin-embedded BM biopsies taken at diagnosis from responders and nonresponders. As shown in Figure 7, C and D, CXCL4 was primarily localized to megakaryocytes, while CXCL7 staining was stronger in an MNC population compatible with a monocytic origin. Importantly, there was increased CXCL4 and CXCL7 staining in BM from nonresponder patients as compared with that in BM from responders, confirming the presence of CXCL4 and CXCL7 proteins in the BM microenvironment, which, like mRNA levels, are increased in DAC-resistant patients.

Previous studies have implicated serum levels of CXCL4 and CXCL7 as potential prognostic markers in MDS (61, 62). To determine whether serum levels of CXCL4 and CXCL7 could potentially serve as biomarkers for DAC response, we quantified the serum concentrations of these chemokines by ELISAs in 35 of 40 CMML patients (Supplemental Figure 6). There was no significant difference in serum CXCL4 or CXCL7 levels between responders and nonresponders. In addition, we found no significant correlation between BM mRNA levels and serum protein levels for these 2 chemokines, indicating that serum levels of these chemokines are not reflective of mRNA expression in the BM and mirroring previous observations documented for other chemokines in the BM and serum of AML patients (63, 64).

CXCL4 and CXCL7 abrogate the effect of DAC on hematopoietic cells. It has been previously reported that both CXCL4 and CXCL7 can reduce the chemosensitivity of BM cells to 5-fluorouracil in vitro (65), and CXCL4 has been implicated in cell cycle arrest (66) and quiescence (67, 68), which might be a mechanism through which it acts to prevent sufficient incorporation of DAC into cells of nonresponders. Therefore, we hypothesized that an overabundance of CXCL4 and CXCL7 in the BM microenvironment acts to overcome the effects of DAC. To test this, we cultured primary human $CD34^+$ cells for 3 days in vitro with CXCL4 (50 ng/ml), CXCL7 (50 ng/ml),

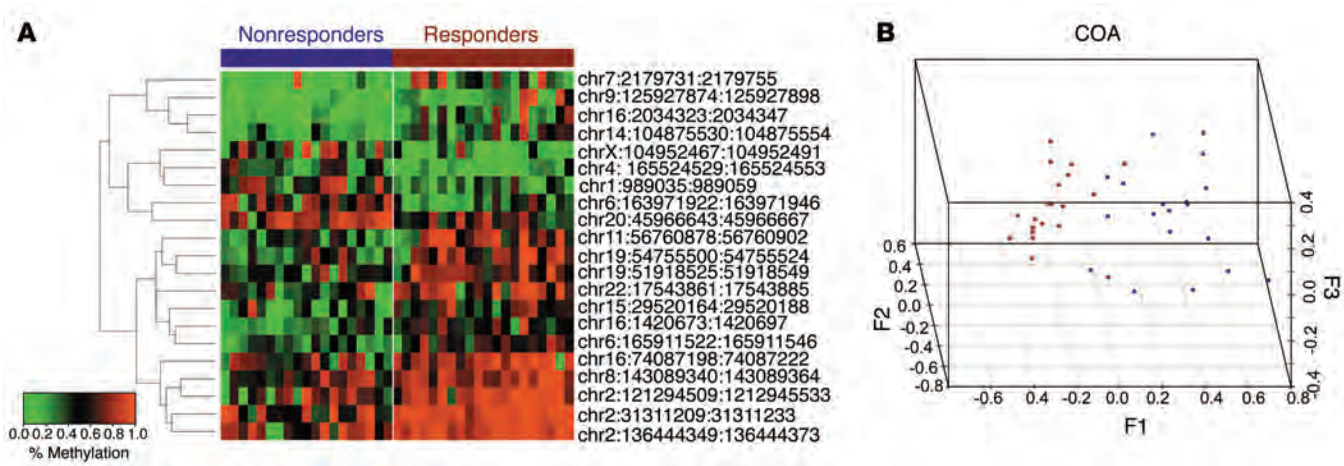


Figure 5. Methylation profiles can be harnessed to classify patients according to DAC response at diagnosis. (**A**) Heatmap of 21 CpG tiles selected as the SVM classifier predictors. DAC-sensitive patients are indicated with the red bar and nonresponders with the blue bar. (**B**) Correspondence analysis (COA) using only the 21 CpG tiles included in the classifier could segregate the majority of the CMML cohort according to DAC response (responders are represented by red dots and nonresponders by blue dots).

or a combination of both chemokines in either the presence or absence of low-dose DAC (10 nM) and then plated them in methylcellulose to test their clonogenic potential. The chemokines and low-dose DAC did not affect cell proliferation during the in vitro liquid culture period (Supplemental Figure 7A). Moreover, as previously reported, low-dose DAC did not reduce cell viability or induce apoptosis after 3 days in culture (Supplemental Figure 7, B and C, and ref. 69). However, 3 days of treatment with 10 nM DAC significantly reduced colony formation. Addition of either CXCL4 or CXCL7 alone did not have a significant impact on DAC-induced colony inhibition. However, concomitant treatment of $CD34^+$ cells with CXCL4 and CXCL7 completely abolished the suppressive effect of DAC on colony formation (Figure 8A).

Finally, we tested the ability of CXCL4 and CXCL7 to induce resistance in primary CMML cells. BM MNCs from diagnostic specimens collected from 3 patients were placed in liquid culture and treated for 72 hours with 10 nM DAC in the presence or absence of 50 ng/ml CXCL4, CXCL7, or a combination of both. Viability was assessed after 72 hours. Unlike normal $CD34^+$ cells, which did not show diminished viability with 10 nM DAC (Supplemental Figure 5B), treatment of primary CMML cells with low-dose DAC led to a significant decrease in viability in all 3 patients ($P < 0.01$). However, concomitant treatment of CMML cells with CXCL4, CXCL7, or their combination abrogated the effect of DAC on all 3 patients (Figure 8B). Combined, these data support the hypothesis that the presence of excess CXCL4 and CXCL7 in the marrow microenvironment contributes to induction of DAC resistance in CMML cells.

Discussion

While DMTis remain the only FDA-approved therapy for the majority of MDS and nonproliferative CMML patients, prognosis following DMTi treatment failure is extremely poor, with median survival for these patients barely reaching 6 months and approximately 50% of patients never even achieving a response in the first place (20, 70). This relatively low rate of therapeutic response is further complicated by the slow kinetics of DMTis, which may take as long as 6 to 12 months to show efficacy, thus committing the majority of patients to receive a drug to which they will ultimately be deemed resistant. Therefore, we set out to study the epigenetic and transcriptional characteristics associated with response to DAC in a cohort of CMML patients in order to identify molecular features that allow risk stratification at the time of diagnosis and, additionally, to explain the mechanisms behind the primary resistance to this agent. To better understand the molecular and mechanistic basis for DMTi response and effectively risk-stratify patients at diagnosis, we performed next-generation sequenc-

Table 3. Prediction performance of the SVM classifier trained on 20 randomly selected samples and applied to the remaining 19 samples in the FISM cohort (accuracy = 94.74%)

Patient ID	Original label	Prediction
1002	NR	NR
0402	*NR*	*R*
0501	NR	NR
0502	R	R
0103	R	R
0105	R	R
0205	NR	NR
0202	R	R
1301	NR	NR
1302	NR	NR
1101	R	R
0204	NR	NR
0507	NR	NR
0802	R	R
0404	NR	NR
0108	R	R
1103	R	R
0901	R	R
0701	R	R

NR, nonresponder; R, responder. Italics indicate an incorrect prediction.

Table 4. Clinical characteristics of the GFM CMML cohort treated with DAC

Clinical characteristics	Responders	Nonresponders	*P* value
Total no. of patients	12	16	
CMML1, no. (%)	2 (17%)	10 (62.5%)	*P* = 0.0235[A]
CMML2, no. (%)	10 (83%)	6 (37.5%)	
Male, no. (%)	9 (75%)	13 (81%)	NS[A]
Female, no. (%)	3 (25%)	3 (19%)	
Median age, yr (range)	72.5 (61–88)	71 (55–85)	NS[B]
Median survival, mo (range)	39 (8–95)	14.5 (5–67)	NS[C]
Median hemoglobin, % (range)	9.1 (6.7–13.3)	9.05 (8–12.2)	NS[A]
Median marrow blasts, % (range)	14 (3–20)	9 (4–19)	NS[D]
Median monocytes, % (range)	23 (2–47)	15.5 (3–34)	NS[D]
Median wbc, % (range)	18.9 (4.9–77.5)	24.95 (4.1–81.7)	NS[A]
Cytogenetics			
Normal	7	11	NS[A]
Abnormal	5	5	

[A]Fisher's exact test; [B]Student's *t* test; [C]log-rank test; [D]Wilcoxon rank-sum test.

ing assays to study both the epigenome and the transcriptome of a uniformly treated cohort of CMML patients who differed in their response to DAC. The use of this improved technology, with extended genomic coverage and better dynamic range, allowed us to detect, for the first time to our knowledge, the presence of DNA methylation and gene expression differences present at the time of diagnosis that distinguish DMTi-sensitive and -resistant patients. The enrichment of these DMRs at distal enhancers, as well as the depletion of promoter-associated DMRs identified in this baseline epigenetic signature, underscores the importance of analyzing DNA methylation changes beyond promoter regions and explains the lack of statistically significant differential methylation observed in previous studies that were confined solely to promoter methylation analysis (12, 27, 30).

Moreover, our observation that the genomic locations predominantly affected by differential DNA methylation are distal regulatory regions adds more data to the strong evidence that emphasizes the critical role of long-range epigenetic gene regulation. Techniques to examine 3D chromatin architecture, such as chromosome conformation capture (3C) (71) and its subsequent iterations 4C (72, 73), 5C (74), and Hi-C (75), have indicated that gene regulation often occurs at very distant locations, in part through DNA looping at distal enhancers. In fact, only a small percentage (~7%) of gene-looping events have been reported to involve the nearest gene transcription start site (50). This argues for the critical role of distal, nonpromoter regulatory regions in controlling gene expression. If the differential methylation at nonpromoter regions does impact the expression of long-range target genes, this may explain why several previous studies have struggled to correlate differential DNA methylation with gene expression changes using nearest-gene annotations (30, 76).

We found that the MAPK pathway was significantly enriched in DMRs, with both gains and losses of methylation in responders and nonresponders within this pathway. These DMRs were localized to both intra- and intergenic genomic regions annotated for 7 genes involved in the MAPK pathway. While in-depth functional analysis of these DMRs will be required in additional experiments that are beyond the scope of our study, our findings support the results by others suggesting the importance of aberrant MAPK pathway signaling in contributing to MDS/MPN (77, 78), as well as to drug resistance and cell cycle progression in leukemic cells (79, 80). Furthermore, while it is known that multiple genes in the MAPK pathway can be mutated in CMML (81), our results indicate that the epigenetic alterations of genes in this pathway may also be present in CMML patients.

While previous reports on MDS and related malignancies have linked the presence of certain mutations — specifically, *TET2* (36–38) and *DNTM3A* (37) — to an increased rate of response to DMTis, we could not find any correlation between the mutational status of these and other genes commonly mutated in CMML and response to DAC in our FISM cohort. This finding is in concordance with those of a previous report on CMML (39), which likewise failed to detect a correlation between response to DAC and mutational status, indicating that the impact of mutational status may be different in CMML patients compared with that in MDS patients or in mixed cohorts consisting of MDS patients as well as patients with other myeloid malignancies, including AML (37, 38) and MDS/MPN (37). Furthermore, the studies demonstrating better *TET2*- and *DNMT3A*-associated responses involved patients treated with AZA alone (38) or cohorts including both AZA- and DAC-treated patients (36, 37), which may also contribute to the differing result obtained in our study on patients who received DAC exclusively.

Conversely, DNA methylation status was indeed different at diagnosis between DAC-sensitive and DAC-resistant patients, and we demonstrate that these differences can risk-stratify patients at the time of diagnosis using an epigenetic classifier that exploits these identified methylation differences. Moreover, the SVM classifier developed in this study performed with 87% accuracy on an independent cohort, even when only a subset of the original features were included and 2 different cell types were used in the training and validation cohorts (BMN MNCs vs. PB monocytes). Thus, while the classifier reported here will require further extensive validation in larger, independent cohorts, the present study demonstrates not only that DNA methylation differences exist between patients with different responses to DAC but that these DNA meth-

Table 5. Summary of the prediction performance of the independent validation cohort (GFM) in 3 scenarios using an increasing number of shared features of the 21 features preselected from the FISM cohort

Number of features used	Correct predictions/ Total patients	Accuracy (%)
16	13/15	87%
14	15/19	79%
6	20/28	71%

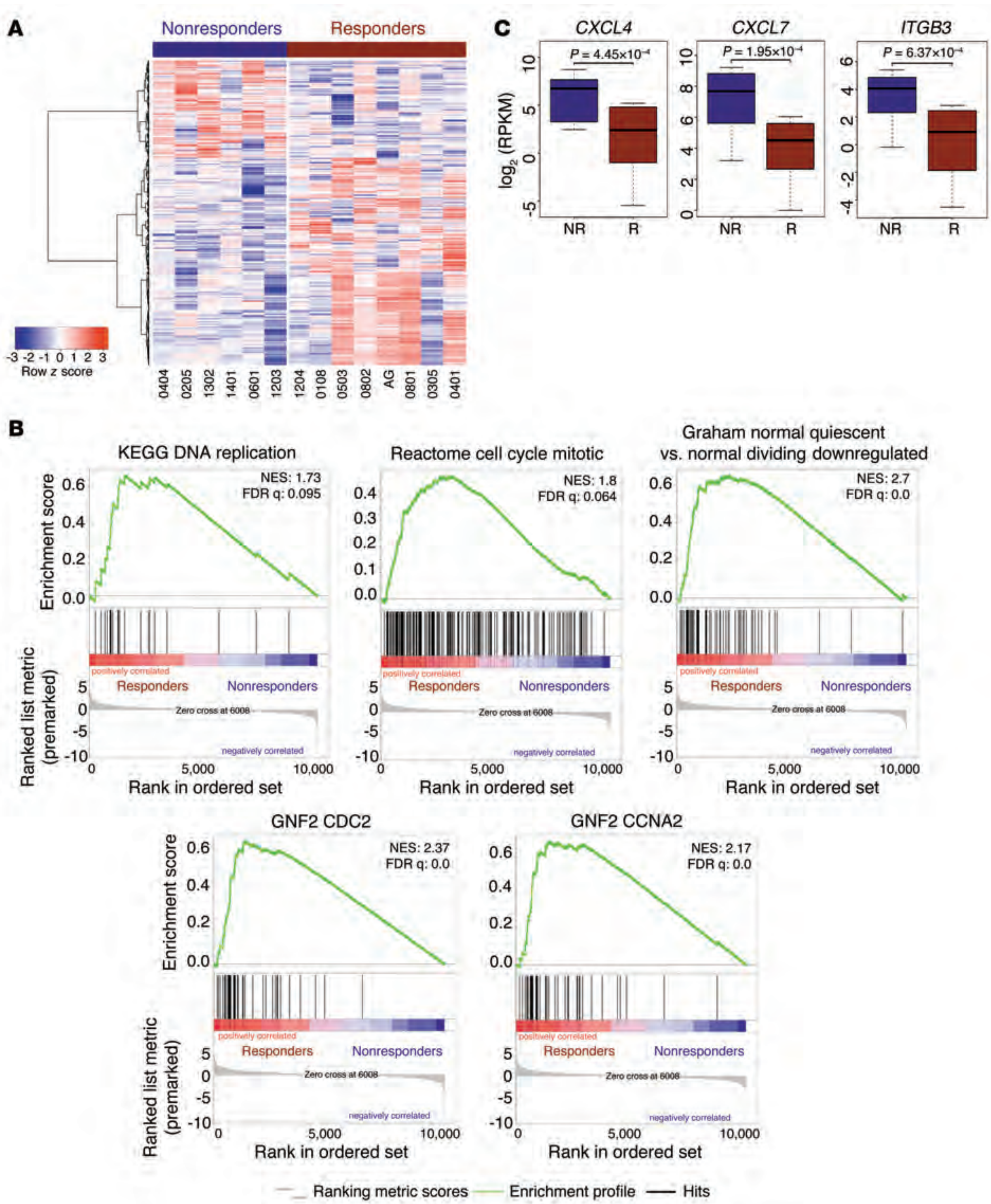


Figure 6. A specific transcriptional program is associated with response to DAC. (**A**) Heatmap illustrates gene expression differences between 8 DAC-sensitive (indicated by the red bar) and 6 DAC-resistant patients (indicated by the blue bar). Genes represented in the heatmap were identified by a GLM likelihood ratio test ($P < 0.05$ and absolute $\log_2$ fold change >1). (**B**) Enrichment plots for GSEA using the expression difference–ranked gene list showing enrichment for cell cycle–related gene sets. NES, normalized enrichment score. (**C**) Box plots showing gene expression differences for *CXCL4*, *CXCL7*, and *ITGB3* (red box plots denote responders; blue box plots denote nonresponders). P values were obtained from a GLM likelihood ratio test.

ylation differences are sufficiently robust to be harnessed for use in the clinic as accurate classifiers. These classifiers have the potential to prevent patients who are unlikely to respond to DAC from receiving prolonged, unwarranted treatments with this drug and instead permit them to be quickly transitioned to alternative therapies.

In addition to epigenetic differences, our study also revealed baseline differences at the transcriptional level that correlated with response to DAC. Analysis of this response-associated signature demonstrated a strong enrichment for gene sets involved in cell cycle regulation among the genes upregulated in DMTi-sensitive

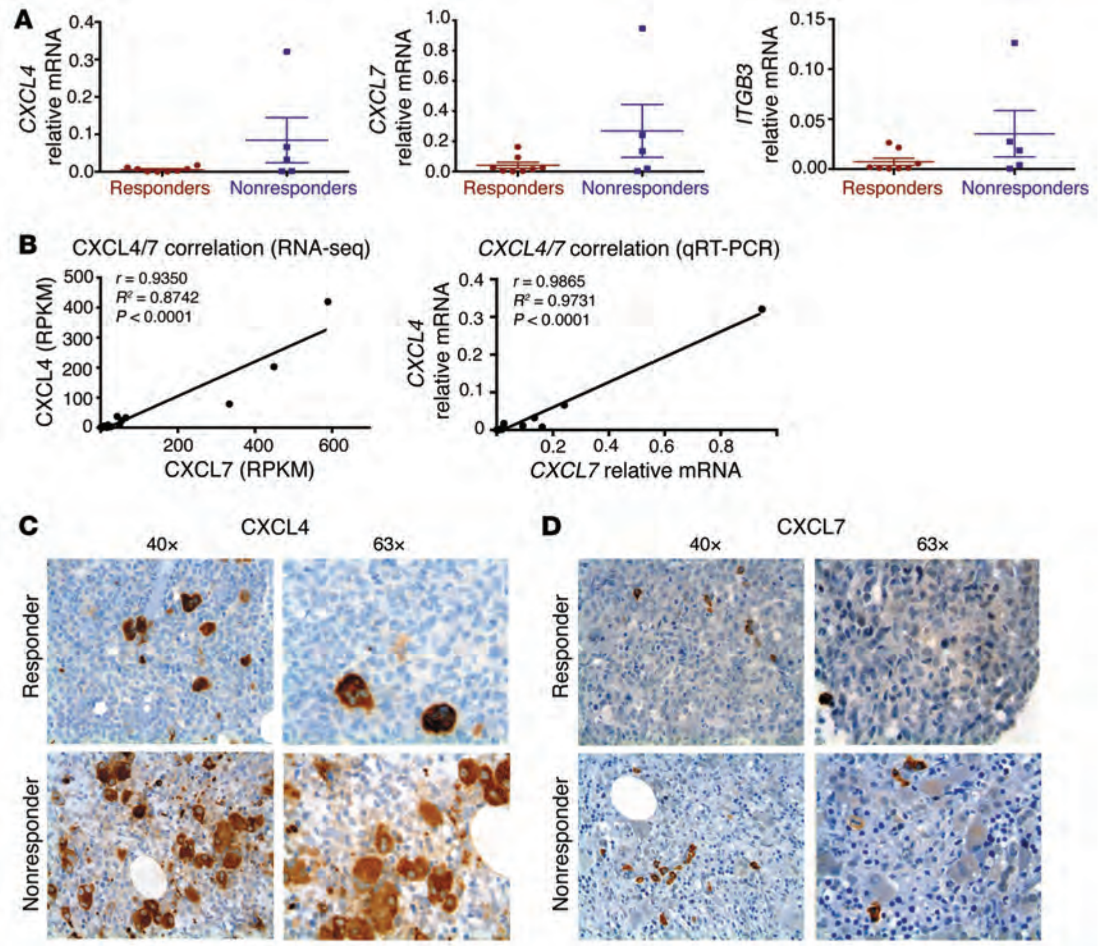


Figure 7. CXCL4 and CXCL7 are upregulated in the BM of nonresponders. (**A**) qRT-PCR showing validation of overexpression of *CXCL4*, *CXCL7*, and *ITGB3* in nonresponders; each point represents the mean of triplicate wells for each patient sample; the line and error bars indicate the group mean and SD, respectively. (**B**) Pearson's correlation analysis of expression levels of CXCL7 and CXCL4 by RNA-seq and qRT-PCR. (**C** and **D**) Representative IHC images for CXCL4 (**C**) and CXCL7 (**D**) in diagnostic BM biopsies in DAC responders and nonresponders. Original magnification, ×40 (**C** and **D**, left panels), ×63 (**C** and **D**, right panels). Representative images from duplicate experiments are shown.

patients. This finding is in line with the need for DAC to be incorporated into the DNA during cell cycle activity in order to exert its effects. By contrast, fewer genes were upregulated in resistant patients. Among these overexpressed genes, we found CXCL4 and CXCL7, two chemokines that have been previously implicated in mediating cell cycle arrest (66), quiescence (67, 68), and reduced chemosensitivity of BM cells to 5-fluorouracil in vitro (65). We therefore focused our efforts on studying the impact of these chemokines on response to DAC. In vitro treatment of both normal CD34$^+$ cells or primary CMML MNCs with CXCL4 and CXCL7 blocked the effect of DAC on these cells, indicating that overexpression of these 2 genes may indeed lead to primary resistance to DAC and opening the possibility for future targeting of the downstream signaling cascades in order to overcome the effect of these chemokines.

Methods

Sample collection and processing

FISM cohort. BM specimens were collected before treatment from 40 patients with CMML. BM MNCs were isolated through Ficoll density centrifugation and viably frozen in 10% DMSO and 90% FBS. Patients with advanced CMML were enrolled in the nonrandomized clinical trial conducted by the FISM (NCT01251627; https://clinicaltrials.gov/) and were given DAC (20 mg/m^2/day i.v.) for 5 days every 28 days for at least 6 cycles prior to being classified as responders or nonresponders, with response defined as HI or better according to IWG 2006 criteria (40). The clinical characteristics of the patients are summarized in Table 1. Genomic DNA and total RNA were isolated using the AllPrep DNA/RNA kit (QIAGEN) according to the manufacturer's instructions.

GFM cohort. The patients were enrolled in the EudraCT 2008-000470-21 GFM trial (NCT01098084; https://www.clinicaltrials.gov/) and received DAC (20 mg/m^2/day i.v.) for 5 days every 28 days for at least 3 cycles. Blood samples were collected using EDTA-containing tubes, mononucleated cells were isolated on Ficoll-Hypaque, and monocytes were enriched using the AutoMacs system (Miltenyi Biotec) through negative selection with microbeads conjugated with antibodies against CD3, CD7, CD16, CD19, CD56, CD123, and glycophorin A, then further enriched by positive selection with microbeads conjugated with a monoclonal mouse anti-human CD14 antibody (Miltenyi Biotec). Genomic DNA was extracted from the monocytes

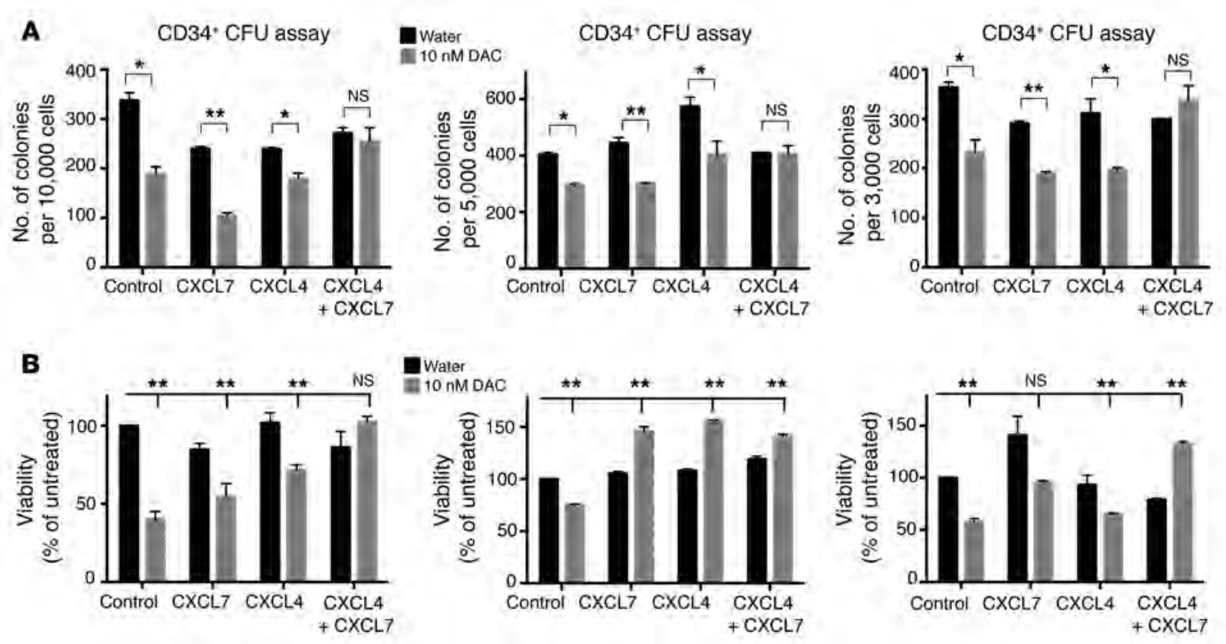


Figure 8. CXCL4 and CXCL7 promote resistance to DAC in CD34⁺ and primary CMML specimens. (**A**) Colony formation was inhibited by DAC but restored with the combination of CXCL4 and CXCL7. CD34⁺ cells were treated with 1 dose of CXCL4, CXCL7, or both (50 ng/ml each) or with vehicle (PBS containing 0.1% BSA) and daily 10-nM doses of DAC for 3 days. After 3 days of in vitro treatment with DAC, cells were plated in methylcellulose and incubated for 12 to 15 days before colonies were counted. Data represent the mean ± SD. Treatment with 10 nM DAC significantly decreased colony formation but failed to do so in the presence of CXCL7 and CXCL4 together. Shown in the 3 panels are the results of 3 independent experiments. Error bars represent the SD. (**B**) CXCL4 and CXCL7 abrogated the effect of DAC on the viability of primary CMML MNCs. CMML MNCs were treated in vitro for 72 hours with 10 nM DAC alone or in the presence of 50 ng/ml CXCL4, CXCL7, or both. Data represent the mean ± SD. Treatment with DAC alone significantly reduced the viability of these cells, but this effect was lost when CXCL4 or CXCL7 was added to the culture. All data represent independent experiments performed in 3 different CMML patients. Error bars represent the SD. *$P < 0.05$ and **$P < 0.01$ by unpaired 2-tailed Student's t test.

using the Norgen Biotek kit (Thorold) kit according to the manufacturer's instructions. The clinical characteristics of the patients are summarized in Table 4.

Mutational sequencing

Target capture. Capture of the target regions (exons plus splice junctions) was carried out using a custom-designed HaloPlex Target Enrichment kit (Agilent Technologies) following the HaloPlex Target Enrichment System-Fast Protocol, version D.5.

Sequencing. DNA (500 ng) from each sample was quantified with a Qubit Fluorometer (Invitrogen) and used in the capture reaction. Each sample had a unique index. Libraries were quantified by Qubit, pooled, and run in an Illumina HiSeq 2500 rapid-run flow cell using the on-board cluster method for paired-end sequencing (2 × 100 bp reads).

Analysis. Sequencing results were demultiplexed and converted to a FASTQ format using Illumina BCL2FASTQ software. The reads were adapter and quality trimmed with Trimmomatic (82) and then aligned to the human genome (UCSC build hg19) using the Burrows-Wheeler Aligner (83). Further local indel realignment and base-quality score recalibration were performed using the Genome Analysis Toolkit (GATK) (84). Single-nucleotide variation and indel calls were generated with the GATK HaplotypeCaller. ANNOVAR (85) was used to annotate variants with functional consequence on genes as well as to identify the presence of these variants in dbSNP 137, the 1000 Genomes Project, ESP6500 (National Heart, Lung, and Blood Institute [NHLBI] GO Exome Sequencing Project), and COSMIC 67.

Genome-wide DNA methylation by ERRBS

High-molecular-weight genomic DNA (25 ng) was used to perform the ERRBS assay as previously described (45) and was sequenced on an Illumina HiSeq 2000. Reads were aligned against a bisulfite-converted human genome (hg18) using Bowtie and Bismark (86). Downstream analysis was performed using R statistical software (version 3.0.3) (87), Bioconductor 2.13 (88), and the MethylSig 0.1.3 package (51). Only genomic regions with coverage ranging from 10 to 500 times were used for the downstream analysis. DMRs were identified by first summarizing the methylation status of the genomic regions into 25-bp tiles and then identifying regions with an absolute methylation difference of 25% or more and an FDR of less than 10%. DMRs were annotated to the RefSeq genes (NCBI) using the following criteria: (a) DMRs overlapping with a gene were annotated to that gene; (b) intergenic DMRs were annotated to all neighboring genes within a 50-kb window; and (c) if no gene was detected within a 50-kb window, then the DMR was annotated to the nearest transcription start site (TSS).

Methylation classifier

SVM (53) was applied using R package e1071 (89) to classify the 2 groups of patients (responders and nonresponders), in which the percentage of methylation of the 25-bp tiles was used as a predictor. The probability mode and sigmoid kernel were used in the SVM function, otherwise the default parameters were applied. We performed 2-step feature selections for the SVM classifier: (a) 25-bp tiles were prefiltered by nominal P values of less than 0.05 and by an absolute methylation difference greater than 20%, calculated using the MethylSig package (51); (b) greedy forward-feature selection was applied on the remaining tiles. Briefly, we assessed and prioritized the predictability of each of the filtered tiles in the SVM model and then sequentially evaluated the combinatorial predictability of the tiles by adding 1 tile from the prioritized tiles to the classifier at a time. The final predictors of the SVM classifier were selected from the set of tiles that could optimally predict patient response. The predictability was assessed on the basis of 10-fold cross-validation. Specifically, we randomly partitioned the 39 samples for which ERRBS libraries were available into 10 complementary subsets, training the SVM model on 9 of the 10 subsets (called the training set) and predicting the classes (responder or nonresponder) on the 1 left-out subset (called the validation set or testing set). To reduce variability, 10 rounds of cross-validation were performed using different partitions, and the validation results were summarized over the rounds. During each round of validation, the probability of each sample being predicted as a responder was recorded, and then the ROC-AUC across 10 rounds was calculated with the R package ROCR (90), and this calculation was used as the assessment of the predictability. Complete code is provided in the Supplemental Methods.

EpiTYPER MassARRAY

Validation of CpG methylation of select genomic regions was performed by MALDI-TOF using EpiTYPER MassARRAY (Sequenom) (49) on bisulfite-converted genomic DNA from a subset of DAC responders and nonresponders. The primers used to amplify these genomic regions and the resultant amplicon sequences are listed in Supplemental Table 6.

RNA-seq

RNA-seq was performed on RNA samples from 14 patients (8 responders and 6 nonresponders) who had high-quality RNA (RNA integrity number >6 as determined by the Agilent 2100 Bioanalyzer). RNA-seq libraries were prepared using the Illumina TruSeq RNA Sample Prep Kit (version 2) according to the manufacturer's instructions. A set of synthetic RNAs from the ERCC (91) at known concentrations were mixed with each of the cDNA libraries. Four separate samples were multiplexed into each lane and sequenced on a HiSeq 2000. The quality of reads obtained was evaluated using FastQC (Babraham Bioinformatics; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The sequenced libraries were aligned to the human genome (hg18) or to the ERCC spike-in reference sequence using TopHat, version 2.0.8 (92), with default parameters.

RNA-seq analysis

HTSeq (0.5.4p5) (93) was used to generate the count matrix with the following parameters: "htseq-count --mode=union --stranded=no" using the following 2 gene transfer format (GTF) annotation files, respectively: (a) the hg18 RefSeq gene GTF file downloaded from the UCSC genome browser for endogenous gene assembly; (b) the ERCC spike-in transcript GTF file downloaded from the official website (http://www.lifetechnologies.com/order/catalog/product/4456740) for ERCC spike-in assembly. The endogenous gene counts were normalized by ERCC spike-in library size, and the differential expression analysis was performed using the edgeR (version 3.4.2; Bioconductor) (94) generalized linear model (GLM). Genes with an absolute $\log_2$ fold change greater than 1 and a P value of less than 0.05 were reported.

qRT-PCR

To validate the RNA-seq results, RNA from selected nonresponder and responder patients was reverse transcribed using the Verso cDNA synthesis kit (Thermo Scientific) with random hexamer primers, according to the manufacturer's instructions. qRT-PCR was performed on the resulting cDNA in triplicate using intron-spanning and -flanking primer sets with Fast SYBR Green Master Mix and the StepOne Plus PCR System (Applied Biosystems) according to the manufacturer's instructions. The primer sequences are listed in Supplemental Table 7.

ELISAs

ELISAs for CXCL4 and CXCL7/NAP2 on serum from the CMML patients were performed using the corresponding ELISA kits (RAB0402 and RAB0135) from Sigma-Aldrich according to the manufacturer's directions. For CXCL4, the serum was diluted 1:500 in the sample dilution buffer provided in the kit.

IHC

For immunostaining, 3-μm-thick formalin-fixed, paraffin-embedded BM sections were deparaffinized in xylenes and hydrated in graded alcohols. Antigen retrieval was performed in EDTA (1 mM, pH 8.0) for two 15-minute cycles at maximum power in a microwave oven, and slides were then incubated with a CXCL4 antibody (1:300, catalog 500-P05; PeptroTech) or a CXCL7 antibody (1:50, catalog orb13423; Biorbyt). Immunostaining was performed with the BenchMark histostainer (Ventana Medical Systems, Roche) using a peroxidase detection kit with DAB substrate according to standard procedures. Sections were then counterstained with hematoxylin.

Cell culture and colony-forming assays

$CD34^+$ cells were isolated from cryopreserved BM MNCs from femoral head specimens using the CD34 MicroBead Isolation Kit (Miltenyi Biotec) according to the manufacturer's instructions. For CMML cells, the cryopreserved BM MNCs were rapidly thawed at 37°C and treated with DNAse to prevent cell clumping. Cells were plated in prestimulation media containing IMDM with 20% BIT (STEMCELL Technologies); IL-6 (20 ng/ml); SCF (100 ng/ml); TPO (100 ng/ml); and FLT3L (10 ng/ml) (PeproTech) and recovered overnight. The following day, either CXCL4 (50 ng/ml; PeproTech); CXCL7 (50 ng/ml; PeproTech); a combination of both chemokines (50 ng/ml each); or vehicle (PBS containing 0.1% BSA) was added as well as freshly prepared DAC (10 nM) (Sigma-Aldrich) or vehicle (water). DAC was replenished daily for a total of 3 days. Live cell numbers and viability were determined by trypan blue exclusion.

For colony assays, an equal number of live, treated $CD34^+$ cells were plated in duplicate in H4435 Enriched MethoCult (STEMCELL Technologies). Colonies were counted after 12 to 15 days.

Apoptosis assays

Apoptosis was assessed using the Tali Apoptosis Kit with annexin V Alexa Fluor 288 and propidium iodide according to the manufacturer's instructions and was measured on a Tali Image-Based Cytometer (all from Life Technologies).

Accession numbers

FISM cohort ERRBS and RNA-seq data are deposited in the NCBI's Gene Expression Omnibus (GEO) database (GEO GSE61163). GFM cohort ERRBS data are also deposited in the GEO database (GEO GSE63787).

Statistics

For the analysis of clinical parameters, Fisher's exact test was used for CMML type and sex; unpaired, 2-tailed Student's t tests were used for clinical parameters with a normal distribution; Wilcoxon signed-rank tests were used when the samples were not normally distributed; and the log-rank test was used for survival. A P value of less than 0.05 was considered significant. Somatic mutations between nonresponders and responders was evaluated using Fisher's exact test, and significance was considered at a P value of less than 0.05. For in vitro cell culture and colony-forming experiments, unpaired, 2-tailed Student's t tests were used for comparisons, and significance was considered at a P value of less than 0.05. For correlation analysis between the RNA-seq and qPCR results, Pearson's correlation was performed, and the r values and P values are indicated in the figures. The ERRBS and RNA-seq analyses were performed using a beta binomial test for differential methylation and a generalized linear model likelihood ratio test for differential gene expression. These methods were implemented through specific algorithms that are described in detail in their respective sections above.

Study approval

The current study was approved by the IRB of the University of Michigan Medical School and the ethics committee of the University of Florence, AOU Careggi-Firenze. The GFM clinical trial EudtraCT (2008-000470-21; https://eudract.ema.europa.eu/) was approved by the ethics committee of the Centre Hospitalier Universitaire de Dijon (Dijon, France). All samples were obtained from patients enrolled in clinical trials, and written, informed consent was obtained from these patients at the time of their enrollment in the study. Samples used in the current study were deidentified prior to use at the University of Michigan.

Acknowledgments

The current study was supported through a Translational Research Program Award from the Leukemia and Lymphoma Society (6016-14, to M.E. Figueroa). K. Meldi was partially supported by a Sass Foundation Postdoctoral Fellowship Award and an NIH T32 Training Grant (T32 CA 9676-22) through the University of Michigan Cancer Biology Program. O. Abdel-Wahab is supported by an NIH K08 Clinical Investigator Award (1K08CA160647-01), a US Department of Defense Postdoctoral Fellow Award in Bone Marrow Failure Research (W81XWH-12-1-0041), the Josie Robertson Investigator Program, and a Damon Runyon Clinical Investigator Award, with support from the Evans Foundation. E. Solary is supported by grants from the French National Cancer Institute and the Ligue Nationale Contre le Cancer. F. Buchi is supported by a Ministero dell'Istruzione dell'Universita e della Ricerca (MIUR)/University of Florence Postdoctoral Fellowship. E. Masala is supported by an FISM Postdoctoral Fellowship. We thank Gianna Baroni and Giulia Raugei for performing IHC.

Address correspondence to: Maria E. Figueroa, Department of Pathology, University of Michigan Medical School, 109 Zina Pitcher Pl., BSRB 2019, Ann Arbor, Michigan 48176, USA. Phone: 734.763.1865; E-mail: marfigue@med.umich.edu. Or to: Valeria Santini, University of Florence AOU Careggi, Hematology, Largo Brambilla 3, 50134 Firenze, Italy. Phone: 39.055.7947296; E-mail: santini@unifi.it.

1. Swerdlow SH, et al., eds. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. 4th ed. Lyon, France: International Agency for Cancer (IARC); 2008.
2. Bennett JM, et al. Proposals for the classification of the myelodysplastic syndromes. *Br J Haematol*. 1982;51(2):189–199.
3. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood*. 2002;100(7):2292–2302.
4. Walter MJ, et al. Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia*. 2011;25(7):1153–1158.
5. Tefferi A, et al. Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN AML. *Leukemia*. 2009;23(7):1343–1345.
6. Abdel-Wahab O, et al. Genetic characterization of TET1, TET2, TET3 alterations in myeloid malignancies. *Blood*. 2009;114(1):144–147.
7. Gelsi-Boyer V, et al. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes chronic myelomonocytic leukaemia. *Br J Haematol*. 2009;145(6):788–800.
8. Khan SN, et al. Multiple mechanisms deregulate EZH2 and histone H3 lysine 27 epigenetic changes in myeloid malignancies. *Leukemia*. 2013;27(6):1301–1309.
9. Ernst T, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet*. 2010;42(8):722–726.
10. Makishima H, et al. Novel homo- and hemizygous mutations in EZH2 in myeloid malignancies. *Leukemia*. 2010;24(10):1799–1804.
11. Nikoloski G, et al. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet*. 2010;42(8):665–667.
12. Figueroa ME, et al. MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation. *Blood*. 2009;114(16):3448–3458.
13. Jiang Y, et al. Aberrant DNA methylation is a dominant mechanism in MDS progression to AML. *Blood*. 2009;113(6):1315–1325.
14. Kantarjian H, et al. Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. *Cancer*. 2006;106(8):1794–1803.
15. Silverman LR, et al. Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the cancer and leukemia group B. *J Clin Oncol*. 2002;20(10):2429–2440.
16. Ghoshal K, et al. 5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, nuclear localization signal. *Mol Cell Biol*. 2005;25(11):4727–4741.
17. Patel K, Dickson J, Din S, Macleod K, Jodrell D, Ramsahoye B. Targeting of 5-aza-2′-deoxycytidine residues by chromatin-associated DNMT1 induces proteasomal degradation of the free enzyme. *Nucleic Acids Res*. 2010;38(13):4313–4324.
18. Palii SS, Van Emburgh BO, Sankpal UT, Brown KD, Robertson KD. DNA methylation inhibitor 5-Aza-2′-deoxycytidine induces reversible genome-wide DNA damage that is distinctly influenced by DNA methyltransferases 1 3B. *Mol Cell Biol*. 2008;28(2):752–771.
19. Cihak A, Weiss JW, Pitot HC. Effects of 5-azacytidine on hepatic polyribosomes maturation of ribosomal RNA. *Acta Biol Med Ger*. 1974;33(5–6):859–865.
20. Griffiths EA, Gore SD. DNA methyltransferase and histone deacetylase inhibitors in the treatment of myelodysplastic syndromes. *Semin Hematol*. 2008;45(1):23–30.
21. Itzykson R, et al. Prognostic factors for response and overall survival in 282 patients with higher-risk myelodysplastic syndromes treated with azacitidine. *Blood*. 2011;117(2):403–411.
22. Gore SD, et al. Combined DNA methyltransferase and histone deacetylase inhibition in the treatment of myeloid neoplasms. *Cancer Res*. 2006;66(12):6361–6369.
23. Daskalakis M, et al. Demethylation of a hypermethylated P15/INK4B gene in patients with myelodysplastic syndrome by 5-Aza-2′-deoxycytidine (decitabine) treatment. *Blood*. 2002;100(8):2957–2964.
24. Kantarjian H, et al. Results of a randomized study of 3 schedules of low-dose decitabine in higher-risk myelodysplastic syndrome chronic myelo-

monocytic leukemia. *Blood*. 2007;109(1):52–57.
25. Mund C, Hackanson B, Stresemann C, Lubbert M, Lyko F. Characterization of DNA demethylation effects induced by 5-Aza-2′-deoxycytidine in patients with myelodysplastic syndrome. *Cancer Res*. 2005;65(16):7086–7090.
26. Blum W, et al. Phase I study of decitabine alone or in combination with valproic acid in acute myeloid leukemia. *J Clin Oncol*. 2007;25(25):3884–3891.
27. Shen L, et al. DNA methylation predicts survival and response to therapy in patients with myelodysplastic syndromes. *J Clin Oncol*. 2010;28(4):605–613.
28. Follo MY, et al. Reduction of phosphoinositide-phospholipase C β1 methylation predicts the responsiveness to azacitidine in high-risk MDS. *Proc Natl Acad Sci U S A*. 2009;106(39):16811–16816.
29. Issa JP, et al. Phase 1 study of low-dose prolonged exposure schedules of the hypomethylating agent 5-aza-2′-deoxycytidine (decitabine) in hematopoietic malignancies. *Blood*. 2004;103(5):1635–1640.
30. Fandy TE, et al. Early epigenetic changes and DNA damage do not predict clinical response in an overlapping schedule of 5-azacytidine and entinostat in patients with myeloid malignancies. *Blood*. 2009;114(13):2764–2773.
31. Bejar R, et al. Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med*. 2011;364(26):2496–2506.
32. Jankowska AM, et al. Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A. *Blood*. 2011;118(14):3932–3941.
33. Kosmider O, et al. TET2 gene mutation is a frequent adverse event in chronic myelomonocytic leukemia. *Haematologica*. 2009;94(12):1676–1681.
34. Yoshida K, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011;478(7367):64–69.
35. Patnaik MM, et al. Spliceosome mutations involving SRSF2, SF3B1, U2AF35 in chronic myelomonocytic leukemia: prevalence, clinical correlates, prognostic relevance. *Am J Hematol*. 2013;88(3):201–206.
36. Bejar R, et al. TET2 mutations predict response to hypomethylating agents in myelodysplastic syndrome patients. *Blood*. 2014;124(17):2705–2712.
37. Traina F, et al. Impact of molecular mutations on treatment response to DNMT inhibitors in myelodysplasia and related neoplasms. *Leukemia*. 2014;28(1):78–87.
38. Itzykson R, et al. Impact of TET2 mutations on response rate to azacitidine in myelodysplastic syndromes low blast count acute myeloid leukemias. *Leukemia*. 2011;25(7):1147–1152.
39. Braun T, et al. Molecular predictors of response to decitabine in advanced chronic myelomonocytic leukemia: a phase 2 trial. *Blood*. 2011;118(14):3824–3831.
40. Cheson BD, et al. Clinical application and proposal for modification of the International Working Group (IWG) response criteria in myelodysplasia. *Blood*. 2006;108(2):419–425.
41. Itzykson R, et al. Prognostic score including gene mutations in chronic myelomonocytic leukemia. *J Clin Oncol*. 2013;31(19):2428–2436.
42. Patnaik MM, et al. Mayo prognostic model for WHO-defined chronic myelomonocytic leukemia: ASXL1 and spliceosome component mutations outcomes. *Leukemia*. 2013;27(7):1504–1510.
43. Meggendorfer M, et al. SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). *Blood*. 2012;120(15):3080–3088.
44. Kohlmann A, et al. Next-generation sequencing technology reveals a characteristic pattern of molecular mutations in 72.8% of chronic myelomonocytic leukemia by detecting frequent alterations in TET2, CBL, RAS, RUNX1. *J Clin Oncol*. 2010;28(24):3858–3865.
45. Akalin A, et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet*. 2012;8(6):e1002781.
46. Figueroa ME, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, impair hematopoietic differentiation. *Cancer Cell*. 2010;18(6):553–567.
47. Figueroa ME, et al. Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. *J Clin Invest*. 2013;123(7):3099–3111.
48. Bullinger L, et al. Quantitative DNA methylation predicts survival in adult acute myeloid leukemia. *Blood*. 2010;115(3):636–642.
49. Ehrich M, et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A*. 2005;102(44):15785–15790.
50. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature*. 2012;489(7414):109–113.
51. Park Y, Figueroa ME, Rozek LS, Sartor MA. MethylSig: a whole genome DNA methylation analysis pipeline. *Bioinformatics*. 2014;30(17):2414–2422.
52. Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol*. 2013;14(3):R21.
53. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273–297.
54. Valencia A, et al. Expression of nucleoside-metabolizing enzymes in myelodysplastic syndromes and modulation of response to azacitidine. *Leukemia*. 2014;28(3):621–628.
55. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550.
56. Graham SM, Vass JK, Holyoake TL, Graham GJ. Transcriptional analysis of quiescent and proliferating CD34+ human hemopoietic cells from normal chronic myeloid leukemia sources. *Stem Cells*. 2007;25(12):3111–3120.
57. Schaffner A, Rhyn P, Schoedon G, Schaer DJ. Regulated expression of platelet factor 4 in human monocytes — role of PARs as a quantitatively important monocyte activation pathway. *J Leukoc Biol*. 2005;78(1):202–209.
58. Pillai MM, Iwata M, Awaya N, Graf L, Torok-Storb B. Monocyte-derived CXCL7 peptides in the marrow microenvironment. *Blood*. 2006;107(9):3520–3526.
59. Bagger FO, et al. HemaExplorer: a database of mRNA expression profiles in normal and malignant haematopoiesis. *Nucleic Acids Res*. 2013;41(Database issue):D1034–D1039.
60. Bagger FO, et al. HemaExplorer: a web server for easy and fast visualization of gene expression in normal and malignant hematopoiesis. *Blood*. 2012;119(26):6394–6395.
61. Aivado M, et al. Serum proteome profiling detects myelodysplastic syndromes and identifies CXC chemokine ligands 4 and 7 as markers for advanced disease. *Proc Natl Acad Sci U S A*. 2007;104(4):1307–1312.
62. Chen C, Bowen DT, Giagounidis AA, Schlegelberger B, Haase S, Wright EG. Identification of disease- and therapy-associated proteome changes in the sera of patients with myelodysplastic syndromes and del(5q). *Leukemia*. 2010;24(11):1875–1884.
63. Reikvam H, Fredly H, Kittang AO, Bruserud O. The possible diagnostic and prognostic use of systemic chemokine profiles in clinical medicine-the experience in acute myeloid leukemia from disease development and diagnosis via conventional chemotherapy to allogeneic stem cell transplantation. *Toxins (Basel)*. 2013;5(2):336–362.
64. Kittang AO, Hatfield K, Sand K, Reikvam H, Bruserud Ø. The chemokine network in acute myelogenous leukemia: molecular mechanisms involved in leukemogenesis and therapeutic implications. *Curr Top Microbiol Immunol*. 2010;341:149–172.
65. Han ZC, et al. Platelet factor 4 and other CXC chemokines support the survival of normal hematopoietic cells reduce the chemosensitivity of cells to cytotoxic agents. *Blood*. 1997;89(7):2328–2335.
66. Gupta SK, Singh JP. Inhibition of endothelial cell proliferation by platelet factor-4 involves a unique action on S phase progression. *J Cell Biol*. 1994;127(4):1121–1127.
67. Dudek AZ, Nesmelova I, Mayo K, Verfaillie CM, Pitchford S, Slungaard A. Platelet factor 4 promotes adhesion of hematopoietic progenitor cells binds IL-8: novel mechanisms for modulation of hematopoiesis. *Blood*. 2003;101(12):4687–4694.
68. Bruns I, et al. Megakaryocytes regulate hematopoietic stem cell quiescence through CXCL4 secretion. *Nat Med*. 2014;20(11):1315–1320.
69. Tsai HC, et al. Transient low doses of DNA-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. *Cancer Cell*. 2012;21(3):430–446.
70. Prebet T, et al. Outcome of high-risk myelodysplastic syndrome after azacitidine treatment failure. *J Clin Oncol*. 2011;29(24):3322–3327.
71. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*. 2002;295(5558):1306–1311.
72. Simonis M, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*. 2006;38(11):1348–1354.
73. Zhao Z, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- interchromosomal interactions. *Nat Genet*. 2006;38(11):1341–1347.
74. Dostie J, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic

elements. *Genome Res*. 2006;16(10):1299–1309.
75. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289–293.
76. Wong YF, Jakt LM, Nishikawa S. Prolonged treatment with DNMT inhibitors induces distinct effects in promoters and gene-bodies. *PLoS One*. 2013;8(8):e71099.
77. Chung E, Hsu CL, Kondo M. Constitutive MAP kinase activation in hematopoietic stem cells induces a myeloproliferative disorder. *PLoS One*. 2011;6(12):e28350.
78. Katsoulidis E, et al. Role of the p38 mitogen-activated protein kinase pathway in cytokine-mediated hematopoietic suppression in myelodysplastic syndromes. *Cancer Res*. 2005;65(19):9029–9037.
79. Abrams SL, et al. The Raf/MEK/ERK pathway can govern drug resistance, apoptosis and sensitivity to targeted therapy. *Cell Cycle*. 2010;9(9):1781–1791.
80. Steelman LS, et al. Dominant roles of the Raf/MEK/ERK pathway in cell cycle progression, prevention of apoptosis and sensitivity to chemotherapeutic drugs. *Cell Cycle*. 2010;9(8):1629–1638.
81. Zhang L, et al. BRAF kinase domain mutations are present in a subset of chronic myelomonocytic leukemia with wild-type RAS. *Am J Hematol*. 2014;89(5):499–504.
82. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120.
83. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
84. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491–498.
85. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
86. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011;27(11):1571–1572.
87. RC Team. *R: A Language And Environment For Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2012.
88. Gentleman RC, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004;5(10):R80.
89. The Comprehensive R Archive Network. e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. TCRAN Web site. http://cran.r-project.org/package=e1071. Accessed February 20, 2015.
90. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. *Bioinformatics*. 2005;21(20):3940–3941.
91. Jiang L, et al. Synthetic spike-in standards for RNA-seq experiments. *Genome Res*. 2011;21(9):1543–1551.
92. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 2013;31(1):46–53.
93. Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166–169.
94. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–140.